

PATENT ABSTRACTS OF JAPAN

JC

(11)Publication number : 2000-357176

(43)Date of publication of application : 26.12.2000

(51)Int.Cl. G06F 17/30
G06F 12/00
G06F 13/00

(21)Application number : 2000-121247

(71)Applicant : INTERNATL BUSINESS MACH CORP
<IBM>

(22)Date of filing : 21.04.2000

(72)Inventor : LEONARDO C MASARAANI

(30)Priority

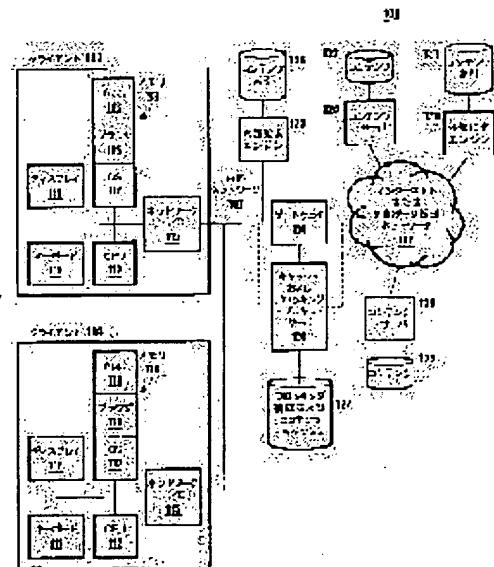
Priority number : 99 302851 Priority date : 30.04.1999 Priority country : US

(54) CONTENTS INDEXING RETRIEVAL SYSTEM AND RETRIEVAL RESULT PROVIDING METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a contents indexing retrieval system which filters and blocks contents by providing a means which implements the same contents blocking policy with a blocking engine by correcting a contents indexing retrieval engine.

SOLUTION: A gateway 124 is used as an interface between a plurality of clients or an internal network 107 and the Internet 106. A proxy server having a cache and blocking proxy server 126 is usually inserted into the connection path from the internal network 107 to the Internet 106 to actualize the contents blocking policy, thereby increasing performance and management. The cache and blocking proxy server 126 can be connected to the gateway 124, or it can be connected directly on parallel to the internal network 107 and an external data transmission network 106.



*** NOTICES ***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

CLAIMS**[Claim(s)]**

[Claim 1]A contents indexing search system which provides search results which are in agreement with filtering of contents, and blocking restrictions, comprising:

A contents indexing search engine containing a database.

Cash and a blocking proxy server containing cash.

An information network combined with said contents indexing search engine.

A means for performing a search inquiry to said contents indexing search engine, and receiving search results from said cash, A blocking engine which combines with said contents indexing search engine, and performs filtering and a blocking policy of contents, A means for correcting said contents indexing search engine and performing the same contents blocking policy as said blocking engine.

[Claim 2]The system according to claim 1 further provided with a means for performing said blocking policy by said contents indexing search engine during a contents indexing phase.

[Claim 3]The system according to claim 1 further provided with a means for performing said blocking policy during an end user search-results display phase.

[Claim 4]The system according to claim 1 further provided with a means to correct said contents indexing search engine in order to build an indexing database by carrying out the index of the cash contents.

[Claim 5]The system according to claim 1 further provided with a means for correcting said contents indexing search engine in order to incorporate said cash and a blocking engine result, when said contents indexing search engine builds an indexing database.

[Claim 6]A contents indexing search engine combined with a database and cash, An information network combined with said contents indexing search engine, In a contents indexing search system which has a blocking engine which performs filtering and blocking restrictions of contents to search results with which an end user is provided via said cash, It is the method of providing search results which are in agreement with filtering of contents, and implementation of blocking -, (a) A step to which arbitrary information site URL which changes a process of said contents indexing search engine, and is in agreement with an exclusion pattern is made to skip, (b) A step which searches only a site or a route contents source which is in agreement with an information site URL list which changes a process of said contents indexing search engine, and is permitted clearly, (c) having a step which performs a filtering policy defined with said cash

and a blocking engine by said contents indexing search engine -- a step of the above (C) -- said filtering policy -- (i), whenever it is a fixed interval or change is detected, A step which reads a filtering rule of contents from said cash and a filtering engine, (ii) A step matched with a user group who generated many indexing database trees, and in whom each tree was provided under a filtering rule of said contents, (iii) Arbitrary information sites which are in agreement with an exclusion pattern, URL, or a step which avoids a display to a user of a document, (iv) A step which displays the document / contents pointer of source origin which is in agreement with an information site URL list permitted clearly, (v) It is set by a step which displays only the information network / contents / document according to a filtering process defined with said cash and said blocking engine on a user, A step which said filtering process is an interval of (aa) regularity, or furthermore reads a filtering rule of said contents from said cash and a filtering engine at a step of said (v) whenever change is detected, (bb) A method defined by a step which displays only search results permitted under said filtering rule to each thing or group to a user.

[Claim 7](d) A step which corrects a contents engine scanning target to a contents cache storage rather than an information site/URL list, (e) A method according to claim 6 of having further a step which traverses URL / contents / document tree of said cash and a blocking engine by API, a database scan, and shared file scan.

[Claim 8](f) A method according to claim 6 of having further a step which corrects said search engine contents scan and an indexing process so that it may be constituted like an end user's browser.

[Claim 9](g) A method according to claim 6 of having further a step which corrects said contents indexing search engine so that said cash may be passed, when building an indexing database.

[Claim 10](h) A method according to claim 6 of having further a step which corrects said contents indexing search engine in order to build an indexing database by searching said cash.

[Claim 11](i) A method according to claim 6 of having further a step which connects said contents indexing search engine to an internal network, and a step which connects the (j) aforementioned contents indexing search engine to internal network operation.

[Claim 12]A method according to claim 6 of having further a step which connects said contents indexing search engine to an external network, and gives compatibility to an organization contents blocking policy.

[Translation done.]

*** NOTICES ***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]**[0001]**

[Field of the Invention]This invention relates to an information retrieval system. Especially this invention relates to the contents indexing search system and method of giving the search results which are in agreement with the policy which performs filtering and blocking of contents which were implemented by the blocking engine.

[0002][Description of the Prior Art]explosive growth of a text and multimedia contents available in the Internet, or other data networks and systems -- an end user -- Seki -- in order to discover prudent information, it depends on the text and the key word retrieving tool increasingly. An end user inputs into a retrieving tool or a search engine the key word showing the information and document for which it is asking. And the list of pointers to the document which a retrieving tool or a search engine refers to the existing indexing database, and is considered to be interesting with the title of a document. It returns with description of several lines which consists of a text which was extracted from the document main part in many cases. Then, an end user navigates some of pointers which have returned for search, or all, and peruses the contents on a actual document or on-line. Generally a search engine indexing database A contents source. An automatic program is started to (for example, the Internet website), By searching a route contents source automatically like the link to a contents tree (in many cases, it moves to other sites), and carrying out the indexing of the information further included in a database for future search, it is built automatically or is built semi-automatically. The search and the indexing which were automated to a large contents source like the website on the Internet are the uniquely practical method of making an index retrieving database.

[0003]It follows on the diversity of available information increasing to an on-line system and a network, a company, an individual, a group, and a network service provider (NSP), A policy and management which restrict the ease of carrying out of the acquisition of such contents for screening the contents it is considered that are things unsuitable for an end user or unnecessary are carried out increasingly. Such a contents managing policy is blocked so that a predetermined on-line service and a network end user may generally receive wholly or in part and unnecessary contents may not arrive. Blocking of contents is performed with other devices inserted between the contents sources generally made into a contents proxy gateway, a data network firewall or an end user, and the purpose. It often prevents that filtering of contents

is realized as some contents cash engines, only the contents of a request for user parent population are held at cash, and cash of the unnecessary contents is carried out. Only by all the users' letting cash pass, they can access network contents. blocking - Contents are carried out when it is what is seen at harmful to a user group or a business use or unsuitable generally, otherwise the specific time of the days. NSP and a company will often depend on the plat form (PICS) for an assessment system or service, for example, Internet-contents selection, in order to judge the conformity of the content site about specific parent population, or a document. An end user may choose a series of blocking policies imposed by themselves by some systems.

[0004]An important problem surfaces to NSP and the data transfer provider who are located between the necessity for the blocking engine which blocks automatically that the necessity for a ***** automatic search engine and some of contents reach an end user eventually in a lot of contents. Especially a problem is lack of the unification with the engine and search engine which carry out filtering and a blocking policy, and cooperation-izing. Lack of unification is produced for several reasons. namely, (a) -- many organizations are adopting and carrying out the policy which performs filtering and blocking - of contents to the site of this organization depending on a search engine, for example, a search engine available on the Internet, or service. (b) There is a tendency for the search engine to have to search and carry out the indexing of as many contents as possible [intentional], and to ask for all the contents positively. On the other hand, in filtering and a blocking engine, selection of the document which stores in cash and is eventually displayed to an end user is planned intentionally.

[0005]Since a role which is intrinsically [between a search engine and a blocking engine] different with execution (implementation) efficiency and the demand which receives highly efficiently exists, cooperation is prevented from unification and cooperation of these two information retrieval functions.

[0006]Based on filtering/blocking policy, eventually the above-mentioned problem in spite of the contents document which cannot be accessed, It becomes clear according to the fact that description and the title of such a contents document are displayed to an end user at the same time it uses service of a search engine. or [that these selves of the title and a short description which are returned by the search engine other than inconvenient / of an end user / or frustration by there being no coherence are dramatically unpleasant] -- or -- otherwise, it is not desirable.

[0007]Therefore, the performance blocking method and the information retrieval system with which the search results been [search results / it] in agreement or adjusted are obtained are called for by few protocols and the performance effect of the grade which can be performed.

[0008]The following are mentioned as contents indexing search and a blocking system.

[0009]U.S. Pat. No. 5,701,469 (Brandli et al.) published on December 23, 1997 removes from a result the preservation search results included accidentally, The contents indexing search system which performs the search-results collection routine which adds the preservation search results excepted accidentally is indicated. Thus, the search results which answered a user's inquiry and were produced are correctly made, even if the contents index used to generate the first search results is not the newest thing.

[0010]U.S. Pat. No. 5,835,722 (Brandshaw et al.) for which it applied on June 27, 1996 and which was published on November 10, 1998, search -- by supervising comprehensively the computer operation for generation of an unsuitable material, or transmission, the terminal for blocking use and transmission of an

unsuitable material is indicated.

Terminal blocking is carried out by it or it is blocked by only surveillance intervention.

[0011]U.S. Pat. No. 5,706,507 (Schloss) for which it applied on July 5, 1995 and which was published on January 6, 1998, In order to block or perceive an unnecessary material, the advisory server operated by the third party who evaluates the contents of the data downloaded from the contents server is indicated.

[0012]U.S. Pat. No. 5,619,648 (Canale et al.) published on April 8, 1997 is indicating E-mail filtering which judges whether a user should be provided with the message of an E-mail based on the model of the relation corresponding to a user.

[0013]. As [become / what only the contents to which any conventional technology was permitted by the blocking policy were returned to the end user as a result of content retrieval, and corresponded with the blocking policy] The contents indexing search system which provides the search results which suit the blocking policy realized by the blocking engine is not indicated.

[0014]

[Problem(s) to be Solved by the Invention]The 1st purpose of this invention is to provide the information retrieval system and method about operation of giving compatibility between a search engine result and a contents blocking policy which were improved.

[0015]The 2nd purpose of this invention is to provide the contents indexing search system and method about operation of giving a blocking policy and the search results to adjust which were improved.

[0016]The 3rd purpose of this invention is to provide the contents indexing search system and method of performing a blocking policy with cash and a filtering engine which were improved.

[0017]The 4th purpose of this invention is to provide the contents indexing search system and method about operation of performing a blocking policy which were improved during a contents indexing phase.

[0018]The 5th purpose of this invention is to provide the contents indexing search system and method about operation of performing a blocking policy which were improved during the phase which displays an end user's search results.

[0019]The 6th purpose of this invention is to provide the contents indexing search system and method about the operation which searches local REPOSITORY of cash and a blocking engine instead of, and carries out the indexing of a final content site and contents server which were improved. [search]

[0020]The 7th purpose of this invention is to provide the contents indexing search system and method about operation used as the composition of passing cash and a filtering engine towards target contents which were improved.

[0021]

[Means for Solving the Problem]These purposes and other purposes, the feature, and an advantage, In order to perform policy control which generally blocks unnecessary contents so that it may be in agreement with an end user's organization filtering and a blocking policy which are performed by embodiment with another search results, It is attained by information retrieval network containing a contents indexing search engine with a cash engine which combines between a search engine and end users, and a database.

[0022]In the 1st example of an embodiment, only contents permitted by blocking policy are added to a search engine indexing database. In the 2nd example of an embodiment, search and a display process of

a search engine are corrected in order to perform a blocking policy. In the 3rd example of an embodiment, by searching contents of a cash engine, operation of a search engine and a target of an indexing automaton process are corrected so that an indexing database may be built. In the 4th example of an embodiment, a scan and an indexing automaton of a search engine are constituted like an end user's browser, and in order to reach [namely,] target contents, they pass cash and a filtering engine.

[0023]

[Embodiment of the Invention]In drawing 1, the information retrieval system 100 has two or more client devices 102,104 connected with the Internet or other distributed data networks 106 via an inside or the controlled network 107. A typical client is a personal computer (PC) with the display 110, the keyboard 111, CPU112, the memory 113, and network connection nature I/O device 115. The home user of PC connected to the business user of PC linked to the network inside a company or a service provider's network as an example of such a client and a network is mentioned, and also when it is which, the final more big Internet is accessed. The browser 116 currently sold with trademarks, such as Netscape Communicator and IBM Web Explorer, is installed in the memory 113 with the standard operating system 117 and the application program 118. The browser 116 is run or performed with the client devices 102 and 104, in order that contents may read or download from the contents server 120 connected with the Internet 106. Each contents server has the database 122 for storing the data which can answer to the contents request from the client 102 and 104 grades. As one gestalt, data is stored as a meeting of the HTML document containing a text and other multimedia contents.

[0024]Generally the gateway 124 is used as an interface of two or more clients or the internal network segment 107, and the Internet 106, as shown in a figure. Usually, the proxy server which has cash and the filtering engine 126 of contents is inserted in the connection path from the internal network 107 to the Internet 106, and performance and management are increased by realizing a contents blocking policy. It can connect with a gateway or cash INGU and a blocking proxy server can be directly connected in parallel with the Internet 107 and the external network 106.

[0025]The client systems 102 and 104 which perform the web browser 116 require contents from the contents server 120 using a HTTP (Hypertext Transfer Protocol) demand, and receive contents by HTTP response. An HTTP request and a response are produced on the TCP/IP socket transmitted on the communication link between a client and a contents server. A user generates a contents request by whether the contents stored in the contents server are required clearly, or the hyperlink anchor which points out the contents stored in the contents server is taken. If it receives, a browser loads the contents which use an HTTP session. The detailed explanation about HTTP is shown in Hypertext Transfer Protocol-HTTP/"1.0" Draft IEFT-HTTP-V10-Spec-0.0 Text 1995 (March 8) of Berners-Lee and others, On these specifications, the whole contents of this literature are used as what makes some of these specifications. The detailed explanation about HTML is shown in "Hypertext Markup Language(HTML)" Draft IEFT.IIIR-HTML-01 of Berners-Lee, and June 1993 (draft out of print), On these specifications, the whole contents of this literature are used as what makes some of these specifications. The detailed explanation about a TCP/IP socket, W. It is shown in "TCP/IP Illustrated and Vol.1 -- The Protocols" of Richard Stevens, Addison-Westlake, 1994 pages 1-20, and 229-262, On these specifications, the whole contents of this literature are used as what makes some of these specifications.

[0026]The user of the client system using the web browser 116, When the most, the conventional search engine servers 130 and 135 and the databases 131 and 136 which were limited in order to trace the place of Internet contents by a retrieval-by-keyword means are accessed, respectively. These fixed engine servers turn into the external search engine server 130 or the internal search engine server 135 to the managed network 107. While they perform the same basic function, by an internal network operator, the search engine server 135 combined with the inside is managed independently, and is rather liked. This invention of process will be realized by the search engine 135 which generally combines with an inside and is managed, or the external search engine 130 which is consistent as service to an organization and gives an organization a contents blocking policy. As a result of the retrieval by keyword turned to the search engine server 130 or 135, an end user, The list of URL which is displayed as a hyperlink anchor to final contents and which text-made an extract and was in agreement will be seen displayed on the web browser 116. The user can choose and follow a link to contents coincidence of one or more using the web browser 116.

[0027]In drawing 2, the table 200 of composition of filtering/blocking a contents sample by a client. Or it is generated by a network / service administrator, and in order to filter the validity of the contents consider that are unsuitable or unnecessary or to limit, it is installed by the proxy server 126. Generally these contents access control methods block with a predetermined on-line service or network all the end users or that receive in part and the contents which are not desirable arrive. A table is installed in cash and the filtering engine 126, and is usually stored in the database 127. As one gestalt, a table receives each user or each user group, A user. Or a group identification descriptor. (ID) The list 205 of keywords to 203 and blocking -, the PICS (Platform For Interconnect Content Selection) rule 207, the black list 209 of URL which should not contact, And even the inside of the white list 211 of only URL which can contact has the line 201 containing the above thing. Explanation of URL is used as "Uniform Resource Locators (URL)" of Berners-Lee and others, and a thing which is shown in RFC 1738 and December 1994 and makes some of these specifications for the whole contents of this literature on these specifications. PICS evaluation is obtained from the PICS rule which is based on the PIC label incorporated into the document which describes URL, and permits or blocks access to URL. The PICS rule is indicated to http: [on the Internet published by W3C etc.]//www.w3.org/TR/REC-PICSRules-971229. Especially a PICS rule is a language for expressing the filtering rule (profile) which permits or blocks access to URL based on the PICS label which describes those URL. A label is generated by using the software tool based on available PICS Technical specification-1.1 by http: [on the Internet]//www.w3.org/PICS/. A software tool is used in order to generate a label by the document which describes specific URL. Or the reader independent of others distributes a label through another server called Ravel Bureau instead of sticking a label into a document. Filtering software will know investigating in Ravel Bureau, in order to find Ravel the same with knowing reading the specific magazine in which consumers criticize an instrument or a private vehicle. Once Ravel is generated, Ravel will be inserted in the HTTP header stream preceded with the contents of the document sent to the web browser as an additional header. Or Ravel can be embedded on a HTML document using a META tag. By this method, Ravel is not a picture, video, or other things, either, and is seen off only by a HTML document. It is available from International Business Machines Corporation (Armonk, NY) in the contents server corresponding to PICS-.

[0028]On the blocking table installed in cash and a blocking engine. It is available in order for some process choices to combine a contents search and a contents blocking engine, and only the contents eventually permitted by the blocking policy are returned to a client as a result of content retrieval. While it is possible to have a different rule to an individual user, it is much more easy to process having a single set of the rule applied to all the users, or that each divides a user into a small number of user group with an original rule. Supposing an individual user or a group is set, it can be identified by using some means by him. As such a means, HTTP basic authentication is used, for example at the time of the start of the client system IP address and browsing section to a user / group ID mapping, Using cookies of HTTPweb for pursuing a user identification is mentioned. [0029]If it returns to drawing 3, the process 300 will perform a blocking policy between contents indexing phases. In Step 302, the contents operation and the indexing automaton process by the search engine 135 are corrected. At Step 304, the filtering rule of contents and the contents from the filtering engine 126 at the usual intervals. Or whenever change is detected, it is read into the search engine server 135 via transmission of an application program interface (API) or a rule definition file. In Step 306, many indexing database trees are generated if needed, and each tree is matched with a user group, as it defined as the filtering rule of contents. For example, one indexing database tree which has attached the PIC filtering rule strict as an object for children is defined, and one indexing database tree with a more nearly free filtering rule as an object for adults is defined.

[0030]In Step 308, the automaton process of a search engine starts a scan, the scan of indexing contents, and an indexing from the list of target servers, while the contents blocking rule is investigated.

[0031]Supposing a white list exists, a search engine will search with Step 310 only the website or route contents source which is clearly in agreement with a permissible site / URL list, or white list.

[0032]In Step 312, if the black list of URL which should be excepted is set up into a rule, website URL which is in agreement with a black list pattern will be excepted anythings.

[0033]In Step 314, it is applied to the site / contents / document currently processed, and a document is excepted as the result, or the PICS rule applied to the user set served with the indexing database tree is included.

[0034]In Step 316, if the list of keywords to except is specified, a document text will be scanned, and if one or more keywords are contained in a list, a document will be removed.

[0035]In Step 318, only when it approves under a filtering rule about the group, a document is added to a suitable indexing database.

[0036]As for the advantage of the process of drawing 3, processing (exclusion) of all the additions is performed in a database indexing phase. There is almost no additional processing needed in user retrieval processing and a presentation phase. Though contents are probably rescanned to a possible change, retrieving operation is performed by the life cycle of a search engine quite more frequently than indexing operation.

[0037]In drawing 4, another process 400 performs a blocking policy between end user search-results display phases. In Step 402, the scan of a search engine and an indexing automaton process do not change, but the single indexing database tree is maintained. In Step 404, in order to apply a blocking policy, search and the display process of a search engine are corrected.

[0038]In Step 406, the filtering rule (passing transmission of an API course or a rule definition) of the

contents from a cash engine is a constant interval, or whenever change is detected, it is read into a search engine.

[0039]In Step 407, processing of the retrieval required which was begun by the user is started to an index database. [0040]In Step 408, the list of all the congruous documents on which a users request is satisfied is created, and it prepares for application of a blocking rule.

[0041]In Step 410, if the white file of URL permitted clearly is specified in a rule, all the documents in agreement which are not contained in a white list will be excepted.

[0042]In Step 412, either the website which is in agreement with an exclusion pattern list (black list) URL or a document is excepted.

[0043]In Step 414, supposing a PICS rule is specified, URL which was not checked under a PICS rule will be excepted also which.

[0044]In Step 416, supposing the keyword list is specified as the rule, each URL which contains one or more keywords in a list in a text will be excepted.

[0045]The remaining subsets of a URL pointer that are satisfied with Step 418 of a blocking rule in accordance with a user's demand are returned for the display to a client.

[0046]The 1st advantage of the process of drawing 4 is that the newest policy can apply without the influence which reconstructs an indexing database to each search. A single indexing database can be used to all the users. The definition which lowers to management of each [without influence] almost, and changes a filtering group by this process is attained.

[0047]The indexing database is built by the process 500 correcting a search engine and searching with drawing 5 the contents of the engine which is carrying out cash of the contents. In Step 501, a search engine scan and an indexing automaton process are corrected. A scan and instead of carrying out an indexing, a process is set up in a final contents source site search the local repository of the contents of cash and a blocking engine. The scanning target of a search engine is corrected to the memory storage which is carrying out cash of the suitable contents more nearly rather than a site/URL list in Step 503. In Step 505, URL / contents / document tree of cash and a blocking engine are traversed via API, database operation, or a share field system operation. In Step 507, since the blocking filtering method to one or more user groups is followed in a local installation, the arbitrary documents found out by cash are added to an indexing database.

[0048]The 1st advantage of the process of drawing 5 is that applying filtering and a blocking rule is performed once [only] with the engine designed such, i.e., cash, and a blocking engine. A scan and an indexing scan are performed on the partial (high performance) copy of target contents rather than a more nearly variable Internet-contents site.

[0049]In drawing 6, since a search engine builds the indexing database of itself, the process 600 corrects a search engine, in order to pass cash and a filtering engine. In Step 601, a search engine scan and an indexing automaton are corrected, and it builds like an end user's browser. That is, since target contents are reached, a HTTP proxy is used and cash and a filtering engine are passed. In Step 603, the search engine automaton has composition which uses the HTTP proxy constituted to suitable cash and a filtering engine. Step 605 -- contents -- a scan -- and a search engine automaton, while carrying out an indexing, It has composition which simulates the end user belonging to one of the user groups so that a user may

receive only the subset containing the site / contents / document permitted by a user group's policy.

[0050]The 1st advantage of the process of drawing 6 is not correcting a search engine substantially.

Contents blocking and filtering are performed with the cash and the blocking engine which were designed and optimized such. Only the contents permitted by the blocking policy reach a search engine for an indexing. Since some of sites / contents by which a scan and an indexing should be made are found out by local cache memory storage, the efficiency and performance of a search engine increase.

[0051]As explained above, content retrieval and a contents blocking engine join together so that only the pointer to the contents eventually permitted by the blocking policy may return to an end user as a result of content retrieval. Many processes are described in order to combine content retrieval and a contents blocking engine. As such a thing, this invention gives filtering of the contents of the result of an end user's content retrieval, and each organization, and the compatibility between blocking policies. This invention can be immediately used without the necessity for change to the existing Internet, other networks, i.e., a data protocol, and standard.

[0052]Although this invention was explained about the Internet (HTTP/Web) environment, the same concept is applied to almost all the data and network environment with which data is searched. The list of possible coincidence will be displayed to the end user who consumes / looks through data in order, if it approves with access or a contents management system. Various change is possible, without deviating from the pneuma and the range of this invention, as it was defined as the claim.

[0053]As a conclusion, the following matters are indicated about the composition of this invention.

(1) The contents indexing search engine which is a contents indexing search system which provides the search results which are in agreement with filtering of contents, and blocking restrictions, and contains a database, The cash and the blocking proxy server containing cash, The information network combined with said contents indexing search engine, The means for performing a search inquiry to said contents indexing search engine, and receiving search results from said cash, The blocking engine which combines with said contents indexing search engine, and performs filtering and the blocking policy of contents, A contents indexing search system provided with the means for correcting said contents indexing search engine and performing the same contents blocking policy as said blocking engine.

(2) A system given in the above (1) further provided with the means for performing said blocking policy by said contents indexing search engine during a contents indexing phase.

(3) A system given in the above (1) further provided with the means for performing said blocking policy during an end user search-results display phase.

(4) A system given in the above (1) further provided with a means to correct said contents indexing search engine in order to build an indexing database by carrying out the index of the cash contents.

(5) A system given in the above (1) further provided with the means for correcting said contents indexing search engine in order to incorporate said cash and a blocking engine result, when said contents indexing search engine builds an indexing database.

(6) The contents indexing search engine combined with a database and cash, The information network combined with said contents indexing search engine, In the contents indexing search system which has a blocking engine which performs filtering and blocking restrictions of contents to the search results with which an end user is provided via said cash, It is the method of providing the search results which are in

agreement with filtering of contents, and implementation of blocking - , (a) The step to which arbitrary information site URL which changes the process of said contents indexing search engine, and is in agreement with an exclusion pattern is made to skip, (b) The step which searches only the site or route contents source which is in agreement with the information site URL list which changes the process of said contents indexing search engine, and is permitted clearly, (c) having a step which performs the filtering policy defined with said cash and a blocking engine by said contents indexing search engine -- the step of the above (C) -- said filtering policy -- (i), whenever it is a fixed interval or change is detected, The step which reads the filtering rule of contents from said cash and a filtering engine, (ii) The step matched with the user group who generated many indexing database trees, and in whom each tree was provided under the filtering rule of said contents, (iii) The arbitrary information sites which are in agreement with an exclusion pattern, URL, or the step which avoids the display to the user of a document, (iv) The step which displays the document / contents pointer of the source origin which is in agreement with the information site URL list permitted clearly, (v) It is set by the step which displays only the information network / contents / document according to the filtering process defined with said cash and said blocking engine on a user, The step which said filtering process is an interval of (aa) regularity, or furthermore reads the filtering rule of said contents from said cash and a filtering engine at the step of said (v) whenever change is detected, (bb) Method defined by the step which displays only the search results permitted under said filtering rule to each thing or group to a user.

(7) The step which corrects (d) contents engine scanning target to a contents cache storage rather than an information site/URL list, (e) A method given in the above (6) which has further a step which traverses URL / contents / document tree of said cash and a blocking engine by API, a database scan, and shared file scan.

(8) Method given in the above (6) which has further a step which corrects said search engine contents scan and an indexing process so that it may be constituted like the (f) end user's browser.

(9) Method given in the above (6) which has further a step which corrects said contents indexing search engine so that said cash may be passed, when building (g) indexing database.

(10) and (h) -- a method given in the above (6) which has further a step which corrects said contents indexing search engine in order to build an indexing database by searching said cash.

(11) -- (i) -- the step which connects said contents indexing search engine to an internal network, and (j) -- a method given in the above (6) which has further a step which connects said contents indexing search engine to internal network operation. (12) A method given in the above (6) which has further a step which connects said contents indexing search engine to an external network, and gives compatibility to an organization contents blocking policy.

[Translation done.]

*** NOTICES ***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

TECHNICAL FIELD

[Field of the Invention]This invention relates to an information retrieval system. Especially this invention relates to the contents indexing search system and method of giving the search results which are in agreement with the policy which performs filtering and blocking of contents which were implemented by the blocking engine.

[0002]

[Translation done.]

*** NOTICES ***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

PRIOR ART

[Description of the Prior Art] explosive growth of a text and multimedia contents available in the Internet, or other data networks and systems -- an end user -- Seki -- in order to discover prudent information, it depends on the text and the key word retrieving tool increasingly. An end user inputs into a retrieving tool or a search engine the key word showing the information and document for which it is asking. And the list of pointers to the document which a retrieving tool or a search engine refers to the existing indexing database, and is considered to be interesting with the title of a document. It returns with description of several lines which consists of a text which was extracted from the document main part in many cases. Then, an end user navigates some of pointers which have returned for search, or all, and peruses the contents on a actual document or on-line. Generally a search engine indexing database A contents source. An automatic program is started to (for example, the Internet website), By searching a route contents source automatically like the link to a contents tree (in many cases, it moves to other sites), and carrying out the indexing of the information further included in a database for future search, it is built automatically or is built semi-automatically. The search and the indexing which were automated to a large contents source like the website on the Internet are the uniquely practical method of making an index retrieving database.

[0003]It follows on the diversity of available information increasing to an on-line system and a network, a company, an individual, a group, and a network service provider (NSP), A policy and management which restrict the ease of carrying out of the acquisition of such contents for screening the contents it is considered that are things unsuitable for an end user or unnecessary are carried out increasingly. Such a contents managing policy is blocked so that a predetermined on-line service and a network end user may generally receive wholly or in part and unnecessary contents may not arrive. Blocking of contents is performed with other devices inserted between the contents sources generally made into a contents proxy gateway, a data network firewall or an end user, and the purpose. It often prevents that filtering of contents is realized as some contents cash engines, only the contents of a request for user parent population are held at cash, and cash of the unnecessary contents is carried out. Only by all the users' letting cash pass, they can access network contents. blocking - Contents are carried out when it is what is seen at harmful to a user group or a business use or unsuitable generally, otherwise the specific time of the days. NSP and a company will often depend on the plat form (PICS) for an assessment system or service, for example, Internet-contents selection, in order to judge the conformity of the content site about specific parent

population, or a document. An end user may choose a series of blocking policies imposed by themselves by some systems.

[0004]An important problem surfaces to NSP and the data transfer provider who are located between the necessity for the blocking engine which blocks automatically that the necessity for a ***** automatic search engine and some of contents reach an end user eventually in a lot of contents. Especially a problem is lack of the unification with the engine and search engine which carry out filtering and a blocking policy, and cooperation-izing. Lack of unification is produced for several reasons. namely, (a) -- many organizations are adopting and carrying out the policy which performs filtering and blocking - of contents to the site of this organization depending on a search engine, for example, a search engine available on the Internet, or service. (b) There is a tendency for the search engine to have to search and carry out the indexing of as many contents as possible [intentional], and to ask for all the contents positively. On the other hand, in filtering and a blocking engine, selection of the document which stores in cash and is eventually displayed to an end user is planned intentionally.

[0005]Since a role which is intrinsically [between a search engine and a blocking engine] different with execution (implementation) efficiency and the demand which receives highly efficiently exists, cooperation is prevented from unification and cooperation of these two information retrieval functions.

[0006]Based on filtering/blocking policy, eventually the above-mentioned problem in spite of the contents document which cannot be accessed, It becomes clear according to the fact that description and the title of such a contents document are displayed to an end user at the same time it uses service of a search engine. or [that these selves of the title and a short description which are returned by the search engine other than inconvenient / of an end user / or frustration by there being no coherence are dramatically unpleasant] -- or -- otherwise, it is not desirable.

[0007]Therefore, the performance blocking method and the information retrieval system with which the search results been [search results / it] in agreement or adjusted are obtained are called for by few protocols and the performance effect of the grade which can be performed.

[0008]The following are mentioned as contents indexing search and a blocking system.

[0009]U.S. Pat. No. 5,701,469 (Brandli et al.) published on December 23, 1997 removes from a result the preservation search results included accidentally, The contents indexing search system which performs the search-results collection routine which adds the preservation search results excepted accidentally is indicated. Thus, the search results which answered a user's inquiry and were produced are correctly made, even if the contents index used to generate the first search results is not the newest thing.

[0010]U.S. Pat. No. 5,835,722 (Brandshaw et al.) for which it applied on June 27, 1996 and which was published on November 10, 1998, search -- by supervising comprehensively the computer operation for generation of an unsuitable material, or transmission, the terminal for blocking use and transmission of an unsuitable material is indicated.

Terminal blocking is carried out by it or it is blocked by only surveillance intervention.

[0011]U.S. Pat. No. 5,706,507 (Schloss) for which it applied on July 5, 1995 and which was published on January 6, 1998, In order to block or perceive an unnecessary material, the advisory server operated by the third party who evaluates the contents of the data downloaded from the contents server is indicated.

[0012]U.S. Pat. No. 5,619,648 (Canale et al.) published on April 8, 1997 is indicating E-mail filtering which judges whether a user should be provided with the message of an E-mail based on the model of the relation corresponding to a user.

[0013]. As [become / what only the contents to which any conventional technology was permitted by the blocking policy were returned to the end user as a result of content retrieval, and corresponded with the blocking policy] The contents indexing search system which provides the search results which suit the blocking policy realized by the blocking engine is not indicated.

[0014]

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

TECHNICAL PROBLEM

[Problem(s) to be Solved by the Invention]The 1st purpose of this invention is to provide the information retrieval system and method about operation of giving compatibility between a search engine result and a contents blocking policy which were improved.

[0015]The 2nd purpose of this invention is to provide the contents indexing search system and method about operation of giving a blocking policy and the search results to adjust which were improved.

[0016]The 3rd purpose of this invention is to provide the contents indexing search system and method of performing a blocking policy with cash and a filtering engine which were improved.

[0017]The 4th purpose of this invention is to provide the contents indexing search system and method about operation of performing a blocking policy which were improved during a contents indexing phase.

[0018]The 5th purpose of this invention is to provide the contents indexing search system and method about operation of performing a blocking policy which were improved during the phase which displays an end user's search results.

[0019]The 6th purpose of this invention is to provide the contents indexing search system and method about the operation which searches local REPOSITORY of cash and a blocking engine instead of, and carries out the indexing of a final content site and contents server which were improved. [search]

[0020]The 7th purpose of this invention is to provide the contents indexing search system and method about operation used as the composition of passing cash and a filtering engine towards target contents which were improved.

[Translation done.]

*** NOTICES ***

JP0 and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

MEANS

[Means for Solving the Problem]These purposes and other purposes, the feature, and an advantage, In order to perform policy control which generally blocks unnecessary contents so that it may be in agreement with an end user's organization filtering and a blocking policy which are performed by embodiment with another search results, It is attained by information retrieval network containing a contents indexing search engine with a cash engine which combines between a search engine and end users, and a database.

[0022]In the 1st example of an embodiment, only contents permitted by blocking policy are added to a search engine indexing database. In the 2nd example of an embodiment, search and a display process of a search engine are corrected in order to perform a blocking policy. In the 3rd example of an embodiment, by searching contents of a cash engine, operation of a search engine and a target of an indexing automaton process are corrected so that an indexing database may be built. In the 4th example of an embodiment, a scan and an indexing automaton of a search engine are constituted like an end user's browser, and in order to reach [namely,] target contents, they pass cash and a filtering engine.

[0023]

[Embodiment of the Invention]In drawing 1, the information retrieval system 100 has two or more client devices 102,104 connected with the Internet or other distributed data networks 106 via an inside or the controlled network 107. A typical client is a personal computer (PC) with the display 110, the keyboard 111, CPU112, the memory 113, and network connection nature I/O device 115. The home user of PC connected to the business user of PC linked to the network inside a company or a service provider's network as an example of such a client and a network is mentioned, and also when it is which, the final more big Internet is accessed. The browser 116 currently sold with trademarks, such as Netscape Communicator and IBM Web Explorer, is installed in the memory 113 with the standard operating system 117 and the application program 118. The browser 116 is run or performed with the client devices 102 and 104, in order that contents may read or download from the contents server 120 connected with the Internet 106. Each contents server has the database 122 for storing the data which can answer to the contents request from the client 102 and 104 grades. As one gestalt, data is stored as a meeting of the HTML document containing a text and other multimedia contents.

[0024]Generally the gateway 124 is used as an interface of two or more clients or the internal network segment 107, and the Internet 106, as shown in a figure. Usually, the proxy server which has cash and the filtering engine 126 of contents is inserted in the connection path from the internal network 107 to the

Internet 106, and performance and management are increased by realizing a contents blocking policy. It can connect with a gateway or cash INGU and a blocking proxy server can be directly connected in parallel with the Internet 107 and the external network 106.

[0025]The client systems 102 and 104 which perform the web browser 116 require contents from the contents server 120 using a HTTP (Hypertext Transfer Protocol) demand, and receive contents by HTTP response. An HTTP request and a response are produced on the TCP/IP socket transmitted on the communication link between a client and a contents server. A user generates a contents request by whether the contents stored in the contents server are required clearly, or the hyperlink anchor which points out the contents stored in the contents server is taken. If it receives, a browser loads the contents which use an HTTP session. The detailed explanation about HTTP is shown in Hypertext Transfer Protocol-HTTP/"1.0" Draft IEFT-HTTP-V10-Spec-0.0 Text 1995 (March 8) of Berners-Lee and others, On these specifications, the whole contents of this literature are used as what makes some of these specifications. The detailed explanation about HTML is shown in "Hypertext Markup Language(HTML)" Draft IEFT.IIIR-HTML-01 of Berners-Lee, and June 1993 (draft out of print), On these specifications, the whole contents of this literature are used as what makes some of these specifications. The detailed explanation about a TCP/IP socket, W. It is shown in "TCP/IP Illustrated and Vol.1 -- The Protocols" of Richard Stevens, Addison-Westlake, 1994 pages 1-20, and 229-262, On these specifications, the whole contents of this literature are used as what makes some of these specifications.

[0026]The user of the client system using the web browser 116, When the most, the conventional search engine servers 130 and 135 and the databases 131 and 136 which were limited in order to trace the place of Internet contents by a retrieval-by-keyword means are accessed, respectively. These fixed engine servers turn into the external search engine server 130 or the internal search engine server 135 to the managed network 107. While they perform the same basic function, by an internal network operator, the search engine server 135 combined with the inside is managed independently, and is rather liked. This invention of process will be realized by the search engine 135 which generally combines with an inside and is managed, or the external search engine 130 which is consistent as service to an organization and gives an organization a contents blocking policy. As a result of the retrieval by keyword turned to the search engine server 130 or 135, an end user, The list of URL which is displayed as a hyperlink anchor to final contents and which text-made an extract and was in agreement will be seen displayed on the web browser 116. The user can choose and follow a link to contents coincidence of one or more using the web browser 116.

[0027]In drawing 2, the table 200 of composition of filtering/blocking a contents sample by a client. Or it is generated by a network / service administrator, and in order to filter the validity of the contents consider that are unsuitable or unnecessary or to limit, it is installed by the proxy server 126. Generally these contents access control methods block with a predetermined on-line service or network all the end users or that receive in part and the contents which are not desirable arrive. A table is installed in cash and the filtering engine 126, and is usually stored in the database 127. As one gestalt, a table receives each user or each user group, A user. Or a group identification descriptor. (ID) The list 205 of keywords to 203 and blocking -, the PICS (Platform For Interconnect Content Selection) rule 207, the black list 209 of URL which should not contact, And even the inside of the white list 211 of only URL which can contact has the line

201 containing the above thing. Explanation of URL is used as "Uniform Resource Locators (URL)" of Berners-Lee and others, and a thing which is shown in RFC 1738 and December 1994 and makes some of these specifications for the whole contents of this literature on these specifications. PICS evaluation is obtained from the PICS rule which is based on the PIC label incorporated into the document which describes URL, and permits or blocks access to URL. The PICS rule is indicated to http: [on the Internet published by W3C etc.]//www.w3.org/TR/REC-PICSRules-971229. Especially a PICS rule is a language for expressing the filtering rule (profile) which permits or blocks access to URL based on the PICS label which describes those URL. A label is generated by using the software tool based on available PICS Technical specification-1.1 by http: [on the Internet]//www.w# .org/PICS/. A software tool is used in order to generate a label by the document which describes specific URL. Or the reader independent of others distributes a label through another server called Ravel Bureau instead of sticking a label into a document. Filtering software will know investigating in Ravel Bureau, in order to find Ravel the same with knowing reading the specific magazine in which consumers criticize an instrument or a private vehicle. Once Ravel is generated, Ravel will be inserted in the HTTP header stream preceded with the contents of the document sent to the web browser as an additional header. Or Ravel can be embedded on a HTML document using a META tag. By this method, Ravel is not a picture, video, or other things, either, and is seen off only by a HTML document. It is available from International Business Machines Corporation (Armonk, NY) in the contents server corresponding to PICS-.

[0028]On the blocking table installed in cash and a blocking engine. It is available in order for some process choices to combine a contents search and a contents blocking engine, and only the contents eventually permitted by the blocking policy are returned to a client as a result of content retrieval. While it is possible to have a different rule to an individual user, it is much more easy to process having a single set of the rule applied to all the users, or that each divides a user into a small number of user group with an original rule. Supposing an individual user or a group is set, it can be identified by using some means by him. As such a means, HTTP basic authentication is used, for example at the time of the start of the client system IP address and browsing section to a user / group ID mapping, Using cookies of HTTPweb for pursuing a user identification is mentioned. [0029]If it returns to drawing 3, the process 300 will perform a blocking policy between contents indexing phases. In Step 302, the contents operation and the indexing automaton process by the search engine 135 are corrected. At Step 304, the filtering rule of contents and the contents from the filtering engine 126 at the usual intervals. Or whenever change is detected, it is read into the search engine server 135 via transmission of an application program interface (API) or a rule definition file. In Step 306, many indexing database trees are generated if needed, and each tree is matched with a user group, as it defined as the filtering rule of contents. For example, one indexing database tree which has attached the PIC filtering rule strict as an object for children is defined, and one indexing database tree with a more nearly free filtering rule as an object for adults is defined.

[0030]In Step 308, the automaton process of a search engine starts a scan, the scan of indexing contents, and an indexing from the list of target servers, while the contents blocking rule is investigated.

[0031]Supposing a white list exists, a search engine will search with Step 310 only the website or route contents source which is clearly in agreement with a permissible site / URL list, or white list.

[0032]In Step 312, if the black list of URL which should be excepted is set up into a rule, website URL

which is in agreement with a black list pattern will be excepted anythings.

[0033]In Step 314, it is applied to the site / contents / document currently processed, and a document is excepted as the result, or the PICS rule applied to the user set served with the indexing database tree is included.

[0034]In Step 316, if the list of keywords to except is specified, a document text will be scanned, and if one or more keywords are contained in a list, a document will be removed.

[0035]In Step 318, only when it approves under a filtering rule about the group, a document is added to a suitable indexing database.

[0036]As for the advantage of the process of drawing 3, processing (exclusion) of all the additions is performed in a database indexing phase. There is almost no additional processing needed in user retrieval processing and a presentation phase. Though contents are probably rescanned to a possible change, retrieving operation is performed by the life cycle of a search engine quite more frequently than indexing operation.

[0037]In drawing 4, another process 400 performs a blocking policy between end user search-results display phases. In Step 402, the scan of a search engine and an indexing automaton process do not change, but the single indexing database tree is maintained. In Step 404, in order to apply a blocking policy, search and the display process of a search engine are corrected.

[0038]In Step 406, the filtering rule (passing transmission of an API course or a rule definition) of the contents from a cash engine is a constant interval, or whenever change is detected, it is read into a search engine.

[0039]In Step 407, processing of the retrieval required which was begun by the user is started to an index database. [0040]In Step 408, the list of all the congruous documents on which a users request is satisfied is created, and it prepares for application of a blocking rule.

[0041]In Step 410, if the white file of URL permitted clearly is specified in a rule, all the documents in agreement which are not contained in a white list will be excepted.

[0042]In Step 412, either the website which is in agreement with an exclusion pattern list (black list) URL or a document is excepted.

[0043]In Step 414, supposing a PICS rule is specified, URL which was not checked under a PICS rule will be excepted also which.

[0044]In Step 416, supposing the keyword list is specified as the rule, each URL which contains one or more keywords in a list in a text will be excepted.

[0045]The remaining subsets of a URL pointer that are satisfied with Step 418 of a blocking rule in accordance with a user's demand are returned for the display to a client.

[0046]The 1st advantage of the process of drawing 4 is that the newest policy can apply without the influence which reconstructs an indexing database to each search. A single indexing database can be used to all the users. The definition which lowers to management of each [without influence] almost, and changes a filtering group by this process is attained.

[0047]The indexing database is built by the process 500 correcting a search engine and searching with drawing 5 the contents of the engine which is carrying out cash of the contents. In Step 501, a search engine scan and an indexing automaton process are corrected. A scan and instead of carrying out an

indexing, a process is set up in a final contents source site search the local repository of the contents of cash and a blocking engine. The scanning target of a search engine is corrected to the memory storage which is carrying out cash of the suitable contents more nearly rather than a site/URL list in Step 503. In Step 505, URL / contents / document tree of cash and a blocking engine are traversed via API, database operation, or a share field system operation. In Step 507, since the blocking filtering method to one or more user groups is followed in a local installation, the arbitrary documents found out by cash are added to an indexing database.

[0048]The 1st advantage of the process of drawing 5 is that applying filtering and a blocking rule is performed once [only] with the engine designed such, i.e., cash, and a blocking engine. A scan and an indexing scan are performed on the partial (high performance) copy of target contents rather than a more nearly variable Internet-contents site.

[0049]In drawing 6, since a search engine builds the indexing database of itself, the process 600 corrects a search engine, in order to pass cash and a filtering engine. In Step 601, a search engine scan and an indexing automaton are corrected, and it builds like an end user's browser. That is, since target contents are reached, a HTTP proxy is used and cash and a filtering engine are passed. In Step 603, the search engine automaton has composition which uses the HTTP proxy constituted to suitable cash and a filtering engine. Step 605 -- contents -- a scan -- and a search engine automaton, while carrying out an indexing, It has composition which simulates the end user belonging to one of the user groups so that a user may receive only the subset containing the site / contents / document permitted by a user group's policy.

[0050]The 1st advantage of the process of drawing 6 is not correcting a search engine substantially. Contents blocking and filtering are performed with the cash and the blocking engine which were designed and optimized such. Only the contents permitted by the blocking policy reach a search engine for an indexing. Since some of sites / contents by which a scan and an indexing should be made are found out by local cache memory storage, the efficiency and performance of a search engine increase.

[0051]As explained above, content retrieval and a contents blocking engine join together so that only the pointer to the contents eventually permitted by the blocking policy may return to an end user as a result of content retrieval. Many processes are described in order to combine content retrieval and a contents blocking engine. As such a thing, this invention gives filtering of the contents of the result of an end user's content retrieval, and each organization; and the compatibility between blocking policies. This invention can be immediately used without the necessity for change to the existing Internet, other networks, i.e., a data protocol, and standard.

[0052]Although this invention was explained about the Internet (HTTP/Web) environment, the same concept is applied to almost all the data and network environment with which data is searched. The list of possible coincidence will be displayed to the end user who consumes / looks through data in order, if it approves with access or a contents management system. Various change is possible, without deviating from the pneuma and the range of this invention, as it was defined as the claim.

[0053]As a conclusion, the following matters are indicated about the composition of this invention.

(1) The contents indexing search engine which is a contents indexing search system which provides the search results which are in agreement with filtering of contents, and blocking restrictions, and contains a database, The cash and the blocking proxy server containing cash, The information network combined with

said contents indexing search engine, The means for performing a search inquiry to said contents indexing search engine, and receiving search results from said cash, The blocking engine which combines with said contents indexing search engine, and performs filtering and the blocking policy of contents, A contents indexing search system provided with the means for correcting said contents indexing search engine and performing the same contents blocking policy as said blocking engine.

(2) A system given in the above (1) further provided with the means for performing said blocking policy by said contents indexing search engine during a contents indexing phase.

(3) A system given in the above (1) further provided with the means for performing said blocking policy during an end user search-results display phase.

(4) A system given in the above (1) further provided with a means to correct said contents indexing search engine in order to build an indexing database by carrying out the index of the cash contents.

(5) A system given in the above (1) further provided with the means for correcting said contents indexing search engine in order to incorporate said cash and a blocking engine result, when said contents indexing search engine builds an indexing database.

(6) The contents indexing search engine combined with a database and cash, The information network combined with said contents indexing search engine, In the contents indexing search system which has a blocking engine which performs filtering and blocking restrictions of contents to the search results with which an end user is provided via said cash, It is the method of providing the search results which are in agreement with filtering of contents, and implementation of blocking -, (a) The step to which arbitrary information site URL which changes the process of said contents indexing search engine, and is in agreement with an exclusion pattern is made to skip, (b) The step which searches only the site or route contents source which is in agreement with the information site URL list which changes the process of said contents indexing search engine, and is permitted clearly, (c) having a step which performs the filtering policy defined with said cash and a blocking engine by said contents indexing search engine -- the step of the above (C) -- said filtering policy -- (i), whenever it is a fixed interval or change is detected, The step which reads the filtering rule of contents from said cash and a filtering engine, (ii) The step matched with the user group who generated many indexing database trees, and in whom each tree was provided under the filtering rule of said contents, (iii) The arbitrary information sites which are in agreement with an exclusion pattern, URL, or the step which avoids the display to the user of a document, (iv) The step which displays the document / contents pointer of the source origin which is in agreement with the information site URL list permitted clearly, (v) It is set by the step which displays only the information network / contents / document according to the filtering process defined with said cash and said blocking engine on a user, The step which said filtering process is an interval of (aa) regularity, or furthermore reads the filtering rule of said contents from said cash and a filtering engine at the step of said (v) whenever change is detected, (bb) Method defined by the step which displays only the search results permitted under said filtering rule to each thing or group to a user.

(7) The step which corrects (d) contents engine scanning target to a contents cache storage rather than an information site/URL list, (e) A method given in the above (6) which has further a step which traverses URL / contents / document tree of said cash and a blocking engine by API, a database scan, and shared file scan.

- (8) Method given in the above (6) which has further a step which corrects said search engine contents scan and an indexing process so that it may be constituted like the (f) end user's browser.
- (9) Method given in the above (6) which has further a step which corrects said contents indexing search engine so that said cash may be passed, when building (g) indexing database.
- (10) and (h) -- a method given in the above (6) which has further a step which corrects said contents indexing search engine in order to build an indexing database by searching said cash.
- (11) -- (i) -- the step which connects said contents indexing search engine to an internal network, and (j) -- a method given in the above (6) which has further a step which connects said contents indexing search engine to internal network operation. (12) A method given in the above (6) which has further a step which connects said contents indexing search engine to an external network, and gives compatibility to an organization contents blocking policy.

[Translation done.]

*** NOTICES ***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1]It is a block diagram for explaining the composition of an information retrieval system based on this invention.

[Drawing 2]It is a table of the contents blocking rule performed in the information retrieval system shown in drawing 1.

[Drawing 3]It is a flow chart for explaining operation of the system of drawing 1 in the 1st example of an embodiment that performs the blocking policy between content retrieval and an indexing phase.

[Drawing 4]It is a flow chart for explaining operation of the system of drawing 1 in the 2nd example of an embodiment that performs a blocking policy during the phase which displays an end user's search results.

[Drawing 5]It is a flow chart for explaining the search engine of drawing 1 in the 3rd example of an embodiment.

[Drawing 6]It is a flow chart for explaining the search engine of drawing 1 in the 4th example of an embodiment.

[Description of Notations]

100 Information retrieval system

102 Client device

104 Client device

106 Internet

107 Network

110 Display

111 Keyboard

112 CPU

113 Memory

115 Network connection nature I/O device

116 Browser

117 Operating system

118 Application program

120 Contents server

122 Database

124 Gateway
126 Cash and the filtering engine of contents
130 Search engine server
131 Database
135 Search engine server
136 Database
200 Table
201 Line
203 Identifier
204 A user or group ID
205 The list of keywords to blocking
207 PICS rule
209 Black list
211 White list
300 Process
400 Process

[Translation done.]

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号
特開2000-357176
(P2000-357176A)

(43)公開日 平成12年12月26日(2000.12.26)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード*(参考)
G 0 6 F 17/30	4 1 4	G 0 6 F 17/30	4 1 4 Z
	1 1 0		1 1 0 F
			1 1 0 E
12/00	5 4 6	12/00	5 4 6 K
13/00	3 5 1	13/00	3 5 1 Z
審査請求 有		請求項の数12 O L (全 13 頁)	最終頁に続く

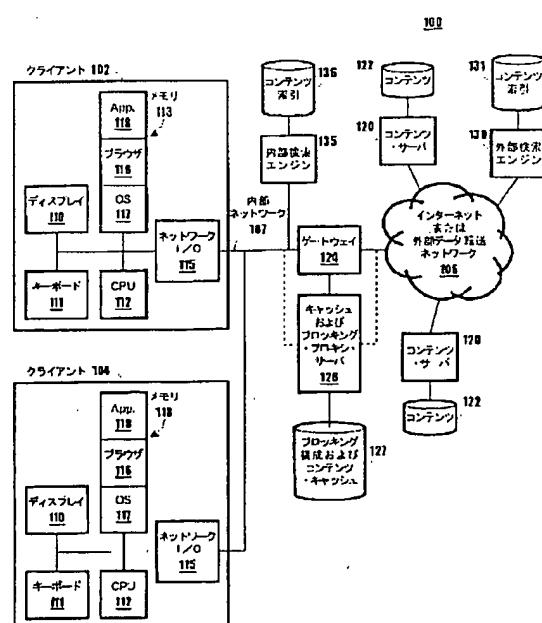
(21)出願番号	特願2000-121247(P2000-121247)	(71)出願人	390009531 インターナショナル・ビジネス・マシーンズ・コーポレーション INTERNATIONAL BUSINESS MACHINES CORPORATION アメリカ合衆国10504、ニューヨーク州 アーモンク (番地なし)
(22)出願日	平成12年4月21日(2000.4.21)	(72)発明者	レオナルド・シー・マサラーニ アメリカ合衆国06880 コネティカット州、 ウェストポート、シグナル・レーン
(31)優先権主張番号	09/302851	(74)代理人	100086243 弁理士 坂口 博 (外1名)
(32)優先日	平成11年4月30日(1999.4.30)		
(33)優先権主張国	米国(US)		

(54)【発明の名称】コンテンツ索引付け検索システム及び検索結果提供方法

(57)【要約】(修正有)

【課題】コンテンツのフィルタリング及びブロッキングを行うコンテンツ索引付け検索システムを提供する。

【解決手段】ユーザは、プロキシ・サーバとキャッシング及びブロッキング・エンジンとしての役割を持つゲートウェイを通して検索エンジンに検索問い合わせを行う。ブロッキング・エンジンは、検索結果に関してコンテンツのフィルタリング及びブロッキングを行い、ユーザのコンテンツ検索結果との間の整合性を与える。又、キャッシング及びフィルタリング・エンジンと同様のコンテンツ・ブロッキング・ポリシーを実行するために、或はキャッシング及びエンジン・コンテンツを検索することで索引データベースを構築するために、或は検索エンジンがそれ自身の索引付けデータベースを構築するときにキャッシング及びフィルタリング・エンジンを通過するために検索エンジンを修正する。



【特許請求の範囲】

【請求項 1】コンテンツのフィルタリング及びブロッキング制限と一致する検索結果を提供するコンテンツ索引付け検索システムであって、データベースを含むコンテンツ索引付け検索エンジンと、キャッシュを含むキャッシュ及びブロッキング・プロキシ・サーバと、前記コンテンツ索引付け検索エンジンに結合した情報ネットワークと、前記コンテンツ索引付け検索エンジンに検索問い合わせを行い、前記キャッシュから検索結果を受け取るための手段と、前記コンテンツ索引付け検索エンジンと結合し、コンテンツのフィルタリング及びブロッキング・ポリシーを実行するブロッキング・エンジンと、前記コンテンツ索引付け検索エンジンを修正して前記ブロッキング・エンジンと同じコンテンツ・ブロッキング・ポリシーを実行するための手段と、を備えるコンテンツ索引付け検索システム。

【請求項 2】コンテンツ索引付けフェーズの間、前記コンテンツ索引付け検索エンジンで前記ブロッキング・ポリシーを実行するための手段をさらに備える請求項 1 に記載のシステム。

【請求項 3】エンド・ユーザ検索結果表示フェーズの間、前記ブロッキング・ポリシーを実行するための手段をさらに備える請求項 1 に記載のシステム。

【請求項 4】キャッシュ・コンテンツを索引することによって索引付けデータベースを構築するために前記コンテンツ索引付け検索エンジンを修正する手段をさらに備える請求項 1 に記載のシステム。

【請求項 5】前記コンテンツ索引付け検索エンジンが索引付けデータベースを構築するときに前記キャッシュ及びブロッキング・エンジン結果を取り込むために前記コンテンツ索引付け検索エンジンを修正するための手段をさらに備える請求項 1 に記載のシステム。

【請求項 6】データベース及びキャッシュに結合したコンテンツ索引付け検索エンジンと、前記コンテンツ索引付け検索エンジンに結合した情報ネットワークと、前記キャッシュを介してエンド・ユーザに提供される検索結果に対してコンテンツのフィルタリング及びブロッキング制限を実行するブロッキング・エンジンとを有するコンテンツ索引付け検索システムにおいて、コンテンツのフィルタリング及びブロッキング・の実施に一致する検索結果を提供する方法であって、

(a) 前記コンテンツ索引付け検索エンジンのプロセスを変えて、除外パターンに一致する任意の情報サイト URL をスキップさせるステップと、

(b) 前記コンテンツ索引付け検索エンジンのプロセスを変えて、明確に許容される情報サイト URL リストに

一致するサイトまたはルート・コンテンツ・ソースのみを検索するステップと、

(c) 前記キャッシュ及びブロッキング・エンジンで定められるフィルタリングポリシーを前記コンテンツ索引付け検索エンジンで実行するステップとを有し、前記

(C) のステップで前記フィルタリングポリシーは、

(i) 一定の間隔で、または変化が検知されるたびに、前記キャッシュ及びフィルタリングエンジンからコンテンツのフィルタリング規則を読み込むステップと、

(ii) 多数索引付けデータベース・ツリーを生成し、各ツリーを前記コンテンツのフィルタリング規則で定められたユーザ・グループと対応付けるステップと、

(iii) 除外パターンに一致する任意の情報サイト、URL、またはドキュメントのユーザに対しての表示を避けるステップと、

(iv) 明確に許容される情報サイト URL リストに一致するソース由来のドキュメント/コンテンツ・ポイントを表示するステップと、

(v) 前記キャッシュ及び前記ブロッキング・エンジンで定められたフィルタリングプロセスに応じる情報ネットワーク/コンテンツ/ドキュメントのみをユーザに表示するステップとによって定められ、さらに前記 (v) のステップで前記フィルタリングプロセスは、

(aa) 一定の間隔で、または変化が検知されるたびに、前記キャッシュ及びフィルタリングエンジンから前記コンテンツのフィルタリング規則を読み込むステップと、

(bb) ユーザに対して個々のものまたはグループに対して前記フィルタリング規則によって許容される検索結果のみを表示するステップによって定められる方法。

【請求項 7】(d) コンテンツ・エンジン走査標的を情報サイト/URL リストよりはむしろコンテンツ・キャッシュ記憶装置に修正するステップと、

(e) API、データベース走査、及び共有ファイル走査によって前記キャッシュ及びブロッキング・エンジンの URL/コンテンツ/ドキュメント・ツリーをトラバースするステップとをさらに有する請求項 6 に記載の方法。

【請求項 8】(f) エンド・ユーザのブラウザと同様にして構成されるように前記検索エンジン・コンテンツ走査及び索引付けプロセスを修正するステップをさらに有する請求項 6 に記載の方法。

【請求項 9】(g) 索引付けデータベースを構築する時に前記キャッシュを通過するように前記コンテンツ索引付け検索エンジンを修正するステップをさらに有する請求項 6 に記載の方法。

【請求項 10】(h) 前記キャッシュを検索することで索引付けデータベースを構築するために前記コンテンツ索引付け検索エンジンを修正するステップをさらに有する請求項 6 に記載の方法。

【請求項11】 (i) 前記コンテンツ索引付け検索エンジンを内部ネットワークに接続するステップと、(j) 前記コンテンツ索引付け検索エンジンを内部ネットワーク操作に接続するステップとをさらに有する請求項6に記載の方法。

【請求項12】 前記コンテンツ索引付け検索エンジンを外部ネットワークに接続して組織コンテンツ・ブロッキング・ポリシーに対して整合性を与えるステップをさらに有する請求項6に記載の方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 この発明は、情報検索システムに関する。特に、本発明はブロッキング・エンジンにインプリメンツされたコンテンツのフィルタリング及びブロッキングを行うポリシーと一致する検索結果を与えるコンテンツ索引付け検索システム及び方法に関する。

【0002】

【従来の技術】 インターネットや他のデータ・ネットワーク及びシステムで利用可能なテキスト及びマルチメディア・コンテンツの爆発的成長によって、エンド・ユーザは、関心ある情報を探し出すためにテキスト及びキー・ワード検索ツールにますます頼っている。エンド・ユーザは、求めている情報及びドキュメントを表しているキー・ワードを検索ツールまたは検索エンジンに入力する。そして、検索ツールまたは検索エンジンは既存の索引付けデータベースで検索を行い、興味あると思われるドキュメントに対するポインタのリストを、ドキュメントのタイトルとともに、また多くの場合ドキュメント本体から抽出されたテキストからなる数行の記述とともに戻してくれる。続いて、エンド・ユーザは検索のために戻ってきたポインタのいくつか、あるいは全てをナビゲートし、実際のドキュメントまたはオンライン上のコンテンツを閲覧する。検索エンジン索引付けデータベースは、一般にコンテンツ・ソース（例えば、インターネット・ウェブ・サイト）に対して自動プログラムを起動し、コンテンツ・ツリー（多くの場合、他のサイトへ移動）へのリンクと同様にルート・コンテンツ・ソースを自動検索し、さらに今後の検索のためにデータベースに含まれる情報を索引付けすることによって、自動的に構築されるか、あるいは半自動的に構築される。インターネット上のウェブ・サイトのような大きいコンテンツ・ソースに対しては、自動化された検索および索引付けは索引検索データベースを作る唯一実用的な方法である。

【0003】 オンライン・システム及びネットワーク、企業、個人、グループ、及びネットワーク・サービス・プロバイダ（N S P）に利用可能な情報の多様性が増加することに伴って、エンド・ユーザにとって不適当あるいは不要なものとみなされるコンテンツをふるいにかけるための、さもなければそのようなコンテンツの入手のしやすさを制限するようなポリシーや管理がますます実

施されている。そのようなコンテンツ管理ポリシーは、一般に所定のオンライン・サービス及びネットワークのエンド・ユーザの全てあるいは一部に対して不要なコンテンツが届かないようにブロッキングする。コンテンツのブロッキングは、一般にコンテンツ・プロキシ・ゲートウェイ、データ・ネットワーク・ファイアウォール、あるいはエンド・ユーザと目的とするコンテンツ・ソースとの間に挿入された他のデバイスで実行される。しばしば、コンテンツのフィルタリングはコンテンツ・キャッシュ・エンジンの一部として実現され、ユーザ母集団のために所望のコンテンツのみがキャッシュに保持されて不要のコンテンツがキャッシュされるのを阻止する。すべてのユーザはキャッシュを通してのみネットワーク・コンテンツにアクセスすることができる。コンテンツは、一般にユーザ・グループまたはビジネス用途にとって有害あるいは不適当なものであること、さもなければ一日のうちの特定の時間に見るものである場合にブロッキング・される。しばしば、N S P及び企業は、特定の母集団に関するコンテンツ・サイトまたはドキュメントの適合性を判断するために、評価システムまたはサービス、例えばインターネット・コンテンツ選択のためのプラットフォーム（P I C S）に頼るであろう。エンド・ユーザはいくつかのシステムで自分自身で課した一連のブロッキング・ポリシーを選択してもよい。

【0004】 自動的に大量のコンテンツを索引付ける自動検索エンジンの必要性と、コンテンツのいくつかがエンド・ユーザに最終的に届くのをブロッキングするブロッキング・エンジンの必要性との間に位置するN S P及びデータ転送プロバイダに対して、重要な問題が浮上する。特に、問題はフィルタリング及びブロッキング・ポリシーを実施するエンジンと検索エンジンとの一体化及び連携化の欠如である。一体化の欠如はいくつかの理由によって生ずる。すなわち、(a) 多くの組織は、検索エンジン、例えばインターネット上で利用可能な検索エンジンに頼る該組織のサイトまたはサービスに対して、コンテンツのフィルタリング及びブロッキング・を行うポリシーを採用して実施している。(b) 検索エンジンは意図的に可能な限り多くのコンテンツを検索して索引付けしなければならず、また全てのコンテンツを積極的に求める傾向がある。一方、フィルタリング及びブロッキング・エンジンでは、キャッシュに格納し、かつエンド・ユーザに対して最終的に表示されるドキュメントの選択が意図的に企てられる。

【0005】 実行（インプリメンテーション）効率及び高性能に対する要求とともに検索エンジンとブロッキングエンジンとの間に本質的に異なる役割が存在することから、これら二つの情報検索機能の一体化及び連係が妨げられる。

【0006】 上記問題は、フィルタリング/ブロッキング・ポリシーにもとづいて最終的にはアクセス不可能

であるコンテンツ・ドキュメントにもかかわらず、検索エンジンのサービスを利用すると同時にそのようなコンテンツ・ドキュメントの記述やタイトルがエンド・ユーザに対して表示されるという事実によって明白となる。首尾一貫性がないことによるエンド・ユーザの不便さやフラストレーションのほかに、検索エンジンによって返されるタイトル及び短い記述は、それら自身が大いに不快であるか、あるいはさもなければ望ましくないものかもしれない。

【0007】したがって、実行可能な程度のわずかなプロトコル及びパフォーマンス効果でパフォーマンス・ブロッキング方式と一致または整合する検索結果が得られる情報検索システムが求められている。

【0008】コンテンツ索引付け検索及びブロッキング・システムとして以下のものが挙げられる。

【0009】1997年12月23日に発行された米国特許第5,701,469号(Brandliら)は誤って含まれてしまった保存検索結果を結果から取り除き、誤って除外された保存検索結果を加える検索結果収集ルーチンを実行するコンテンツ索引付け検索システムを開示している。このようにして、ユーザの問い合わせに応答して生じた検索結果は、最初の検索結果を生成するのに使われたコンテンツ索引が最新のものでなくとも正確に作られる。

【0010】1996年6月27日に出願され1998年11月10日に発行された米国特許第5,835,722号(Brandshawら)は、検索不適な材料の生成または転送のためのコンピュータ操作を包括的に監視することによって、不適当な材料の使用及び転送をブロッキングするためのターミナルを開示しており、それによってターミナルブロッキングされるか、あるいは監視介入によってのみブロッキングされる。

【0011】1995年7月5日に出願され1998年1月6日に発行された米国特許第5,706,507号(Schloss)は、不要な材料をブロッキングまたは感知するために、コンテンツ・サーバからダウンロードされたデータのコンテンツを評価するサード・パーティによって操作されるアドバイザリ・サーバを開示している。

【0012】1997年4月8日に発行された米国特許第5,619,648号(Canaleら)は、ユーザ対応関係のモデルにもとづいて電子メールのメッセージがユーザに提供されるべきかどうかを判断する電子メール・フィルタリングを開示している。

【0013】いずれの従来技術も、ブロッキング・ポリシーによって許可されたコンテンツのみがコンテンツ検索の結果としてエンド・ユーザに戻されて検索結果はブロッキング・ポリシーに一致したものとなるよう、ブロッキング・エンジンに実現されたブロッキング・ポリシーに適合する検索結果を提供するコンテンツ索引付け検索システムを開示していない。

【0014】

【発明が解決しようとする課題】本発明の第1の目的は、検索エンジン結果とコンテンツ・ブロッキング・ポリシーとの間に整合性を与える操作に関する改良された情報検索システム及び方法を提供することである。

【0015】本発明の第2の目的は、ブロッキング・ポリシーと整合する検索結果を与える操作に関する改良されたコンテンツ索引付け検索システム及び方法を提供することである。

【0016】本発明の第3の目的は、キャッシュ及びフィルタリングエンジンでブロッキング・ポリシーを実行する改良されたコンテンツ索引付け検索システム及び方法を提供することである。

【0017】本発明の第4の目的は、コンテンツ索引付けフェーズの間、ブロッキング・ポリシーを実行する操作に関する改良されたコンテンツ索引付け検索システム及び方法を提供することである。

【0018】本発明の第5の目的は、エンド・ユーザの検索結果を表示するフェーズの間、ブロッキング・ポリシーを実行する操作に関する改良されたコンテンツ索引付け検索システム及び方法を提供することである。

【0019】本発明の第6の目的は、検索の代わりにキャッシュ及びブロッキング・エンジンのローカル・レポートリを検索し、最終的なコンテンツ・サイト及びコンテンツ・サーバを索引付けする操作に関する改良されたコンテンツ索引付け検索システム及び方法を提供することである。

【0020】本発明の第7の目的は、標的コンテンツに向けてキャッシュ及びフィルタリングエンジンを通過する構成となった操作に関する改良されたコンテンツ索引付け検索システム及び方法を提供することである。

【0021】

【課題を解決するための手段】これらの目的及び他の目的、特徴及び利点は、検索結果が別の実施形態で実行されるエンド・ユーザの組織フィルタリング及びブロッキング・ポリシーと一致するように一般に不要なコンテンツをブロッキングする管理ポリシーを実行するために、検索エンジンとエンド・ユーザとの間を結合するキャッシュ・エンジンとデータベースとを持つコンテンツ索引付け検索エンジンを含む情報検索ネットワークで達成される。

【0022】第1の実施形態例では、ブロッキング・ポリシーに許容されるコンテンツのみが検索エンジン索引付けデータベースに追加される。第2の実施形態例では、検索エンジンの検索及び表示プロセスはブロッキング・ポリシーを実行するために修正される。第3の実施形態例では、検索エンジンの操作及び索引付けオートマトン・プロセスの標的是キャッシュ・エンジンのコンテンツを検索することによって索引付けデータベースを構築するように修正される。第4の実施形態例では、検索エンジンの走査及び索引付けオートマトンはエンド・ユ

ーザのブラウザと同様にして構成されるもので、すなわち標的コンテンツに到達するためにキャッシュ及びフィルタリング・エンジンを通過する。

【0023】

【発明の実施の形態】図1において、情報検索システム100は内部または制御されたネットワーク107を介してインターネットまたは他の分散型データ・ネットワーク106と接続している複数のクライアント・デバイス102, 104を有する。典型的なクライアントは、ディスプレイ110、キーボード111、CPU112、メモリ113及びネットワーク接続性入出力デバイス115を持つパーソナル・コンピュータ(PC)である。そのようなクライアント及びネットワークの例として、会社内部のネットワークに接続したPCのビジネス・ユーザやサービス・プロバイダのネットワークに接続したPCのホーム・ユーザが挙げられ、どちらの場合も最終的にはより大きなインターネットに接続する。Netscape Communicator, IBM Web Explorer等の商標で販売されているブラウザ116は、標準的なオペレーティング・システム117及びアプリケーション・プログラム118とともにメモリ113にインストールされる。ブラウザ116は、インターネット106につながったコンテンツ・サーバ120からコンテンツの読み込みまたはダウンロードするために、クライアント・デバイス102、104でランまたは実行される。各コンテンツ・サーバは、クライアント102、104等からのコンテンツ要求に対して応答可能なデータを格納するためのデータベース122を有する。一つの形態として、データはテキスト及び他のマルチメディア・コンテンツを含むHTMLドキュメントの集まりとして格納される。

【0024】ゲートウェイ124は、図に示すように、複数のクライアントまたは内部ネットワーク・セグメント107とインターネット106とのインターフェースとして一般に使用される。通常、キャッシュ及びコンテンツのフィルタリング・エンジン126を有するプロキシ・サーバは内部ネットワーク107からインターネット106への接続パスに挿入され、コンテンツ・ブロッキング・ポリシーを実現することでパフォーマンス及び管理を増す。キャッシュイング及びブロッキング・プロキシ・サーバはゲートウェイに接続することができ、あるいはインターネット107と外部ネットワーク106とに平行して直接的に接続することができる。

【0025】ウェブ・ブラウザ116を実行するクライアント・システム102、104は、HTTP(Hypertext Transfer Protocol)要求を用いてコンテンツ・サーバ120からコンテンツを要求し、HTTP応答でコンテンツを受け取る。HTTP要求及び応答は、クライアントとコンテンツ・サーバとの間の通信リンク上で伝達されるTCP/IPソケット上で生ずる。ユーザは、コンテンツ・サーバに格納されたコンテンツを明確に要

求するか、あるいはコンテンツ・サーバに格納されたコンテンツを指すハイパーリンク・アンカーをとるかによって、コンテンツ要求を生成する。受信すると、ブラウザはHTTPセッションを使用するそのコンテンツをロードする。HTTPに関する詳細な説明は、Berners-Leeらの“Hypertext Transfer Protocol-HTTP/1.0”Draft I EFT-HTTP-V10-Spec-0.0 Text 1995(March 8)に示されており、本明細書ではこの文献の内容全体を本明細書の一部をなすものとして援用する。HTMLに関する詳細な説明は、Berners-Leeの“Hypertext Markup Language(HTML)”Draft I EFT.IIIR-HTML-01, June 1993(絶版ドラフト)に示されており、本明細書ではこの文献の内容全体を本明細書の一部をなすものとして援用する。TCP/IPソケットに関する詳細な説明は、W. Richard Stevensの“TCP/IP Illustrated, Vol. 1-The Protocols”, Addison-Westlake, 1994 pages 1-20, 229-262に示されており、本明細書ではこの文献の内容全体を本明細書の一部をなすものとして援用する。

【0026】Webブラウザ116を用いるクライアント・システムのユーザは、たいていの場合、キーワード検索手段によってインターネット・コンテンツの場所を突き止めるために限定された従来の検索エンジン・サーバ130、135及びデータベース131、136にそれぞれアクセスする。これらの固定エンジン・サーバは、管理されたネットワーク107に対して外部検索エンジン・サーバ130または内部検索エンジン・サーバ135となる。それらが同一の基本機能を実行する間に、内部に結合した検索エンジン・サーバ135は内部ネットワーク・オペレータによって独立して管理され、どちらかといえば好まれる。この方法の発明は、一般に内部に結合して管理されるサーチ・エンジン135によって、あるいは組織に対するサービスとして首尾一貫して組織にコンテンツ・ブロッキング・ポリシーを与える外部検索エンジン130によって、実現されるであろう。検索エンジン・サーバ130または135に向けられたキーワード検索の結果として、エンド・ユーザは、最終的なコンテンツに対するハイパーリンク・アンカーとして表示されるテキスト抜粋及び一致したURLのリストがWebブラウザ116に表示されるのを見るであろう。ユーザは、Webブラウザ116を用いて一つ以上のコンテンツ一致に対してリンクを選択し、たどることができる。

【0027】図2では、コンテンツ・サンプルをフィルタリング/ブロッキングする構成のテーブル200がクライアントによって、またはネットワーク/サービス・アドミニストレータによって生成され、不適当または不要と見なされるコンテンツの有効性をフィルタリングするか、さもなければ限定するためにプロキシ・サーバ126でインストールされる。これらのコンテンツ・アクセス管理方式は、一般に所定のオンライン・サービスま

たはネットワークでエンド・ユーザの全てまたは一部に対して、望ましくないコンテンツが届くのをブロッキングする。テーブルはキャッシュ及びフィルタリングエンジン126にインストールされ、通常はデータベース127に格納される。一つの形態として、テーブルは各ユーザまたは各ユーザ・グループに対して、ユーザまたはグループ識別子(1D)203、ブロッキング・に対するキーワードのリスト205、PICS(Platform For Interconnect Content Selection)規則207、接触すべきではないURLのブラック・リスト209、及び接触することができるURLのみのホワイト・リスト211のうち一つ以上のものを含む行201を有する。URLの説明は、Berners-Leeらの"Uniform Resource Locators(URL)"、RFC 1738, December 1994に示されており、本明細書ではこの文献の内容全体を本明細書の一部をなすものとして援用する。PICS評価は、URLを記述するドキュメントに取り込まれたPICSラベルにもとづいてURLへのアクセスを許可またはブロッキングするPICS規則から得られる。PICS規則は、W3C等で発行されたインターネット上のhttp://www.w3.org/TR/REC-PICSRules-971229に記載されている。特に、PICS規則はそれらのURLを記述するPICSラベルにもとづいてURLに対するアクセスを許可またはブロッキングするフィルタリング規則(プロファイル)を表すための言語である。ラベルは、インターネット上のhttp://www.w3.org/PICS/で入手可能なPICS Technicals pecification-1.1にもとづくソフトウェア・ツールを用いることで生成される。ソフトウェア・ツールは、特定のURLを記述するドキュメントでラベルを生成するために使われる。あるいは、ドキュメントの中にラベルを貼る代わりに、他に依存しない読み取りプログラムがラベル・ビューロと呼ばれる別のサーバを通してラベルを配る。フィルタリング・ソフトウェアは、消費者が器具または自家用車を論評する特定の雑誌を読むことを知るのと同じように、ラベルを見付けるためにラベル・ビューロにおいて調べることを知るであろう。ひとたびラベルが生成されると、ラベルはWebブラウザに送られたドキュメントのコンテンツに先行するHTTPヘッダ・ストリームに追加のヘッダとして挿入される。あるいは、METAタグを用いてHTMLドキュメントにラベルを埋め込むことができる。この方法で、ラベルは画像、ビデオ、あるいは他のものでもなくHTMLドキュメントだけで送られる。PICS-対応のコンテンツ・サーバがInternational Business Machines Corporation (Armonk, NY)から入手可能である。

【0028】キャッシュ及びブロッキング・エンジンにインストールされたブロッキング・テーブルによって、いくつかのプロセス選択肢がコンテンツ・サーチとコンテンツ・ブロッキング・エンジンとを結合させるために利用可能であり、ブロッキング・ポリシーによって最終

的に許可されるコンテンツのみがコンテンツ検索の結果としてクライアントに戻される。各個人ユーザに対して異なる規則を持つことが可能であると同時に、全てのユーザに対して適用される規則の單一セットを持つこと、あるいは各々が独自の規則を持つ少數のユーザ・グループにユーザを分けることがよりいっそう処理しやすい。個人ユーザまたはグループは、もし定められるならば、いくつかの手段を使うことによって識別されることが可能である。そのような手段としては、例えばユーザ/グループIDマッピングに対するクライアント・システムIPアドレス、ブラウジング・セクションの開始時にHTTP基本認証を用いること、ユーザ識別を追跡するためのHTTPwebのcookiesを使用することが挙げられる。

【0029】図3に戻ると、プロセス300はコンテンツ索引付けフェーズの間にブロッキング・ポリシーを実行する。ステップ302では、検索エンジン135によるコンテンツ操作及び索引付けオートマトン・プロセスが修正される。ステップ304ではコンテンツ及びフィルタリングエンジン126からのコンテンツのフィルタリング規則を、通常の間隔で、または変化が検出されたたびに、アプリケーション・プログラム・インターフェース(API)または規則定義ファイルの転送を介して検索エンジン・サーバ135に読み込まれる。ステップ306では、多数の索引付けデータベース・ツリーが必要に応じて生成され、各ツリーはコンテンツのフィルタリング規則で定義されたようにユーザ・グループと対応づけられる。例えば、子供用として厳格なPICSフィルタリング規則を付けている一つの索引付けデータベース・ツリーが定義され、大人用としてもっと自由なフィルタリング規則を持つ一つの索引付けデータベース・ツリーが定義される。

【0030】ステップ308では、サーチ・エンジンのオートマトン・プロセスは、コンテンツ・ブロッキング規則が調べられている間に走査及び索引付けコンテンツの走査及び索引付けを標的のサーバのリストから開始する。

【0031】ステップ310では、もしホワイト・リストが存在するならば、検索エンジンは明確に許容可能なサイト/URLリストまたはホワイト・リストに一致するWebサイトまたはルート・コンテンツ・ソースのみを検索する。

【0032】ステップ312では、もし除外すべきURLのブラック・リストが規則の中に設定されるならば、ブラック・リスト・パターンに一致するWebサイトURLはいかなるものでも除外される。

【0033】ステップ314では、索引付けデータベース・ツリーによってサービスされているユーザ・セットに適用されるPICS規則は、処理されているサイト/コンテンツ/ドキュメントに適用され、その結果としてドキュメントが除外されるか、もしくは含まれる。

【0034】ステップ316では、除外するキーワードのリストが特定されるならば、ドキュメント・テキストを走査して一つ以上のキーワードがリストに含まれるならばドキュメントを除去する。

【0035】ステップ318では、そのグループに関してフィルタリング規則によって許容される場合のみ、ドキュメントは適当な索引付けデータベースに加えられる。

【0036】図3のプロセスの利点は、全ての追加の（排除）処理はデータベース索引付けフェーズで実行される。ユーザ検索処理及びプレゼンテーション・フェーズで必要とされる追加の処理はほとんどない。たぶん、可能性のある変化に対してコンテンツが再走査されるとしても、検索操作は検索エンジンのライフ・サイクルで索引付け操作よりもかなり頻繁に行われる。

【0037】図4では、エンド・ユーザ検索結果表示フェーズの間に別のプロセス400がブロッキング・ポリシーを実行する。ステップ402では、検索エンジンの走査及び索引付けオートマトン・プロセスは変化せず、単一の索引付けデータベース・ツリーが維持されている。ステップ404では、ブロッキング・ポリシーを適用するためサーチ・エンジンの検索及び表示プロセスが修正される。

【0038】ステップ406では、キャッシング・エンジンからのコンテンツのフィルタリング規則（API経由または規則定義の転送を介して）は一定間隔で、あるいは変化が検出されるたびに検索エンジンに読み込まれる。

【0039】ステップ407では、ユーザによって始められた検索要求の処理が索引データベースに対して開始される。

【0040】ステップ408では、ユーザ要求を満足させる全ての一一致しているドキュメントのリストを作成し、ブロッキング規則の適用のために準備する。

【0041】ステップ410では、明らかに許容されるURLのホワイト・ファイルが規則のなかで特定されるならば、ホワイト・リストに含まれない全ての一一致するドキュメントは除外される。

【0042】ステップ412では、除外パターン・リスト（ブラック・リスト）に一致するWebサイト、URL、またはドキュメントのいずれも除外される。

【0043】ステップ414では、もしPICS規則が特定されるならば、PICS規則によって確認されなかったURLはどれでも除外される。

【0044】ステップ416では、もしキーワード・リストが規則に特定されているならば、リストにある一つ以上のキーワードをテキストに含むURLはいずれも除外される。

【0045】ステップ418では、ユーザの要求に一致し、ブロッキング規則を満足するURLポイントの残り

のサブセットがクライアントに対する表示のために戻される。

【0046】図4のプロセスの第1の利点は、索引付けデータベースを再構築する影響なしに最新のポリシーが各検索に適用できることである。単一の索引付けデータベースはすべてのユーザに対して使用することができる。このプロセスによってほとんど影響なしに個々の管理まで下げるフィルタリンググループを変更する定義が可能となる。

【0047】図5では、プロセス500は検索エンジンを修正して、コンテンツをキャッシングしているエンジンのコンテンツを検索することでその索引付けデータベースを構築する。ステップ501では、検索エンジン走査及び索引付けオートマトン・プロセスが修正される。最終的なコンテンツ・ソース・サイトを走査及び索引付けする代わりに、プロセスはキャッシング及びブロッキング・エンジンのコンテンツのローカル・リポジトリを検索するように設定される。ステップ503では、検索エンジンの走査標的は、サイト/URLリストよりはむしろ適当なコンテンツをキャッシングしている記憶装置に修正される。ステップ505では、キャッシング及びブロッキング・エンジンのURL/コンテンツ/ドキュメント・ツリーはAPI、データベース操作または共有ファイルド・システム操作を介してトラバースされる。ステップ507では、ローカル・インストレーションにおいて1つ以上のユーザ・グループに対するブロッキング・フィルタリング方式に従うので、キャッシングに見いだされる任意のドキュメントが索引付けデータベースに加えられる。

【0048】図5のプロセスの第1の利点はフィルタリング及びブロッキング規則を適用することがどのように設計されたエンジン、すなわちキャッシング及びブロッキング・エンジンによってたった一回行われることである。走査及び索引付け走査は、より可変的なインターネット・コンテンツ・サイトよりはむしろ標的的コンテンツの局所（高性能）コピー上で実行される。

【0049】図6において、検索エンジンはそれ自体の索引付けデータベースを構築するので、プロセス600はキャッシング及びフィルタリングエンジンを通過するために検索エンジンを修正する。ステップ601では、検索エンジン走査及び索引付けオートマトンを修正してエンド・ユーザのブラウザと同様に構築する。すなわち、標的的コンテンツに達するためにHTTPプロキシを用いてキャッシング及びフィルタリングエンジンを通過する。ステップ603では、検索エンジン・オートマトンは適当なキャッシング及びフィルタリングエンジンに対して構成されているHTTPプロキシを使用する構成となっている。ステップ605では、コンテンツを走査及び索引付けする一方で、検索エンジン・オートマトンは、ユーザ・グループのポリシーによって許容されるサイト/コ

ンテンツ／ドキュメントを含むサブセットのみをユーザが受け取るように、ユーザ・グループの一つに属するエンド・ユーザをシミュレートする構成となっている。

【0050】図6のプロセスの第1の利点は、検索エンジンを実質的に修正しないことである。コンテンツ・ブロッキング及びフィルタリングは、そのように設計及び最適化されたキャッシュ及びブロッキング・エンジンによって実行される。ブロッキング・ポリシーによって許容されたコンテンツのみが索引付けのために検索エンジンに達する。走査及び索引付けがなされるべきサイト／コンテンツのいくつかはローカル・キャッシュ記憶装置に見いだされるため、検索エンジンの効率及び性能は増加する。

【0051】以上説明したように、ブロッキング・ポリシーによって最終的に許容されるコンテンツに対するポイントのみがコンテンツ検索の結果としてエンド・ユーザに戻るよう、コンテンツ検索及びコンテンツ・ブロッキング・エンジンが結合する。コンテンツ検索及びコンテンツ・ブロッキング・エンジンを結合させるために多数のプロセスが記述される。そのようなものとして、本発明はエンド・ユーザのコンテンツ検索の結果と個々の組織のコンテンツのフィルタリング及びブロッキング・ポリシーとの間の整合性を与える。本発明は既存のインターネット及び他のネットワーク、すなわちデータ・プロトコル及び標準に対して、変更の必要性無しにただちに利用することが可能である。

【0052】本発明をインターネット（HTTP/Web）環境について説明したが、同様の概念はデータが検索されるほとんどのデータ及びネットワーク環境に適用される。可能性のある一致のリストは、アクセスまたはコンテンツ管理方式によって許容されるならばデータを順に消費／一覧するエンド・ユーザに対して表示される。特許請求の範囲に定義したように本発明の精神及び範囲から逸脱することなく様々な変更が可能である。

【0053】まとめとして、本発明の構成に関して以下の事項を開示する。

(1) コンテンツのフィルタリング及びブロッキング制限と一致する検索結果を提供するコンテンツ索引付け検索システムであって、データベースを含むコンテンツ索引付け検索エンジンと、キャッシュを含むキャッシュ及びブロッキング・プロキシ・サーバと、前記コンテンツ索引付け検索エンジンに結合した情報ネットワークと、前記コンテンツ索引付け検索エンジンに検索問い合わせを行い、前記キャッシュから検索結果を受け取るための手段と、前記コンテンツ索引付け検索エンジンと結合し、コンテンツのフィルタリング及びブロッキング・ポリシーを実行するブロッキング・エンジンと、前記コンテンツ索引付け検索エンジンを修正して前記ブロッキング・エンジンと同じコンテンツ・ブロッキング・ポリシーを実行するための手段と、を備えるコンテンツ索引付

け検索システム。

(2) コンテンツ索引付けフェーズの間、前記コンテンツ索引付け検索エンジンで前記ブロッキング・ポリシーを実行するための手段をさらに備える上記(1)に記載のシステム。

(3) エンド・ユーザ検索結果表示フェーズの間、前記ブロッキング・ポリシーを実行するための手段をさらに備える上記(1)に記載のシステム。

(4) キャッシュ・コンテンツを索引することによって索引付けデータベースを構築するために前記コンテンツ索引付け検索エンジンを修正する手段をさらに備える上記(1)に記載のシステム。

(5) 前記コンテンツ索引付け検索エンジンが索引付けデータベースを構築するときに前記キャッシュ及びブロッキング・エンジン結果を取り込むために前記コンテンツ索引付け検索エンジンを修正するための手段をさらに備える上記(1)に記載のシステム。

(6) データベース及びキャッシュに結合したコンテンツ索引付け検索エンジンと、前記コンテンツ索引付け検索エンジンに結合した情報ネットワークと、前記キャッシュを介してエンド・ユーザに提供される検索結果に対してコンテンツのフィルタリング及びブロッキング制限を実行するブロッキング・エンジンとを有するコンテンツ索引付け検索システムにおいて、コンテンツのフィルタリング及びブロッキング・の実施に一致する検索結果を提供する方法であって、(a) 前記コンテンツ索引付け検索エンジンのプロセスを変えて、除外パターンに一致する任意の情報サイトURLをスキップさせるステップと、(b) 前記コンテンツ索引付け検索エンジンのプロセスを変えて、明確に許容される情報サイトURLリストに一致するサイトまたはルート・コンテンツ・ソースのみを検索するステップと、(c) 前記キャッシュ及びブロッキング・エンジンで定められるフィルタリングポリシーを前記コンテンツ索引付け検索エンジンで実行するステップとを有し、前記(c)のステップで前記フィルタリングポリシーは、(i) 一定の間隔で、または変化が検知されるたびに、前記キャッシュ及びフィルタリングエンジンからコンテンツのフィルタリング規則を読み込むステップと、(ii) 多数索引付けデータベース・ツリーを生成し、各ツリーを前記コンテンツのフィルタリング規則で定められたユーザ・グループと対応付けるステップと、(iii) 除外パターンに一致する任意の情報サイト、URL、またはドキュメントのユーザに対しての表示を避けるステップと、(iv) 明確に許容される情報サイトURLリストに一致するソース由来のドキュメント／コンテンツ・ポインタを表示するステップと、(v) 前記キャッシュ及び前記ブロッキング・エンジンで定められたフィルタリングプロセスに応じる情報ネットワーク／コンテンツ／ドキュメントのみをユーザに表示するステップとによって定められ、さらに前

記(v)のステップで前記フィルタリングプロセスは、(a a)一定の間隔で、または変化が検知されるたびに、前記キャッシュ及びフィルタリングエンジンから前記コンテンツのフィルタリング規則を読み込むステップと、(b b)ユーザーに対して個々のものまたはグループに対して前記フィルタリング規則によって許容される検索結果のみを表示するステップによって定められる方法。

(7) (d) コンテンツ・エンジン走査標的を情報サイト/URLリストよりはむしろコンテンツ・キャッシュ記憶装置に修正するステップと、(e) API、データベース走査、及び共有ファイル走査によって前記キャッシュ及びブロッキング・エンジンのURL/コンテンツ/ドキュメント・ツリーをトラバースするステップとをさらに有する上記(6)に記載の方法。

(8) (f) エンド・ユーザーのブラウザと同様にして構成されるように前記検索エンジン・コンテンツ走査及び索引付けプロセスを修正するステップをさらに有する上記(6)に記載の方法。

(9) (g) 索引付けデータベースを構築する時に前記キャッシュを通過するように前記コンテンツ索引付け検索エンジンを修正するステップをさらに有する上記(6)に記載の方法。

(10) (h) 前記キャッシュを検索することで索引付けデータベースを構築するために前記コンテンツ索引付け検索エンジンを修正するステップをさらに有する上記(6)に記載の方法。

(11) (i) 前記コンテンツ索引付け検索エンジンを内部ネットワークに接続するステップと、(j) 前記コンテンツ索引付け検索エンジンを内部ネットワーク操作に接続するステップとをさらに有する上記(6)に記載の方法。(12) 前記コンテンツ索引付け検索エンジンを外部ネットワークに接続して組織コンテンツ・ブロッキング・ポリシーに対して整合性を与えるステップをさらに有する上記(6)に記載の方法。

【図面の簡単な説明】

【図1】本発明にもとづく情報検索システムの構成を説明するためのブロック図である。

【図2】図1に示す情報検索システムにおいて実行されるコンテンツ・ブロッキング規則の表である。

【図3】コンテンツ検索及び索引付けフェーズの間のブロッキング・ポリシーを実行する第1の実施形態例における図1のシステムの操作を説明するためのフローチャ

ートである。

【図4】エンド・ユーザーの検索結果を表示するフェーズの間、ブロッキング・ポリシーを実行する第2の実施形態例における図1のシステムの操作を説明するためのフローチャートである。

【図5】第3の実施形態例における図1の検索エンジンを説明するためのフローチャートである。

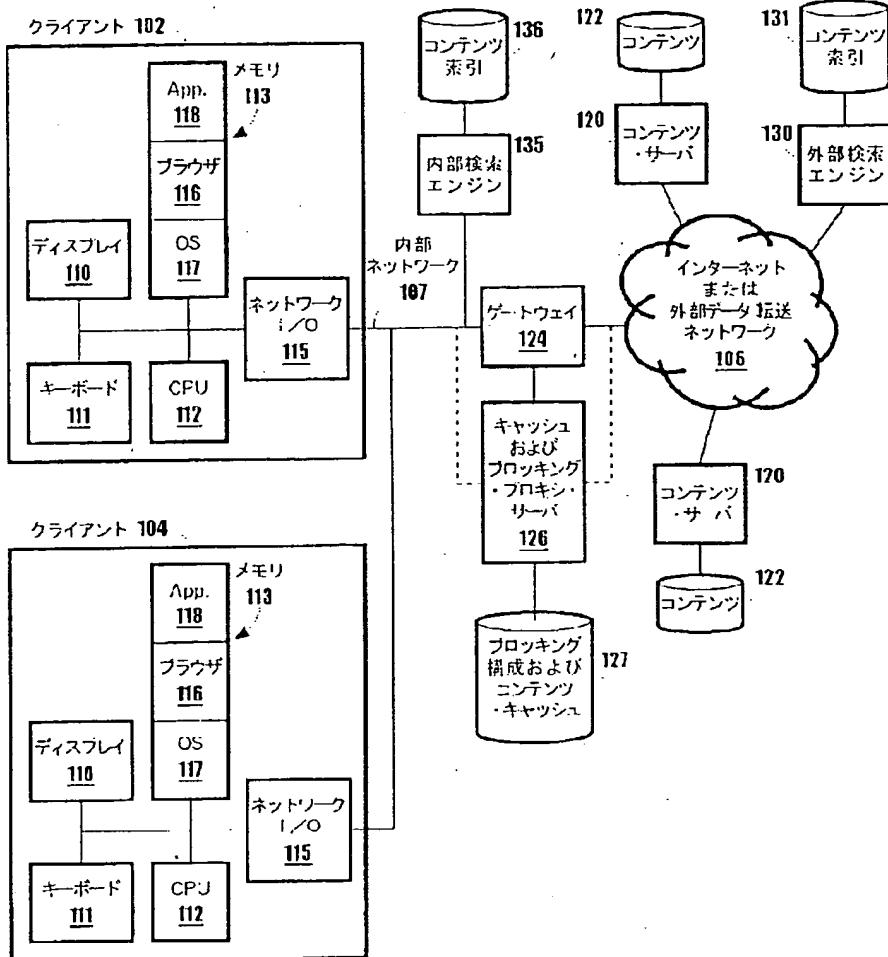
【図6】第4の実施形態例における図1の検索エンジンを説明するためのフローチャートである。

【符号の説明】

1 0 0	情報検索システム
1 0 2	クライアント・デバイス
1 0 4	クライアント・デバイス
1 0 6	インターネット
1 0 7	ネットワーク
1 1 0	ディスプレイ
1 1 1	キーボード
1 1 2	C P U
1 1 3	メモリ
1 1 5	ネットワーク接続性入出力デバイス
1 1 6	ブラウザ
1 1 7	オペレーティング・システム
1 1 8	アプリケーション・プログラム
1 2 0	コンテンツ・サーバ
1 2 2	データベース
1 2 4	ゲートウェイ
1 2 6	キャッシュ及びコンテンツのフィルタリング・エンジン
1 3 0	検索エンジン・サーバ
1 3 1	データベース
1 3 5	検索エンジン・サーバ
1 3 6	データベース
2 0 0	テーブル
2 0 1	行
2 0 3	識別子
2 0 4	ユーザまたはグループ ID
2 0 5	ブロッキングに対するキーワードのリスト
2 0 7	P I C S 規則
2 0 9	ブラック・リスト
2 1 1	ホワイト・リスト
3 0 0	プロセス
4 0 0	プロセス

【図1】

100



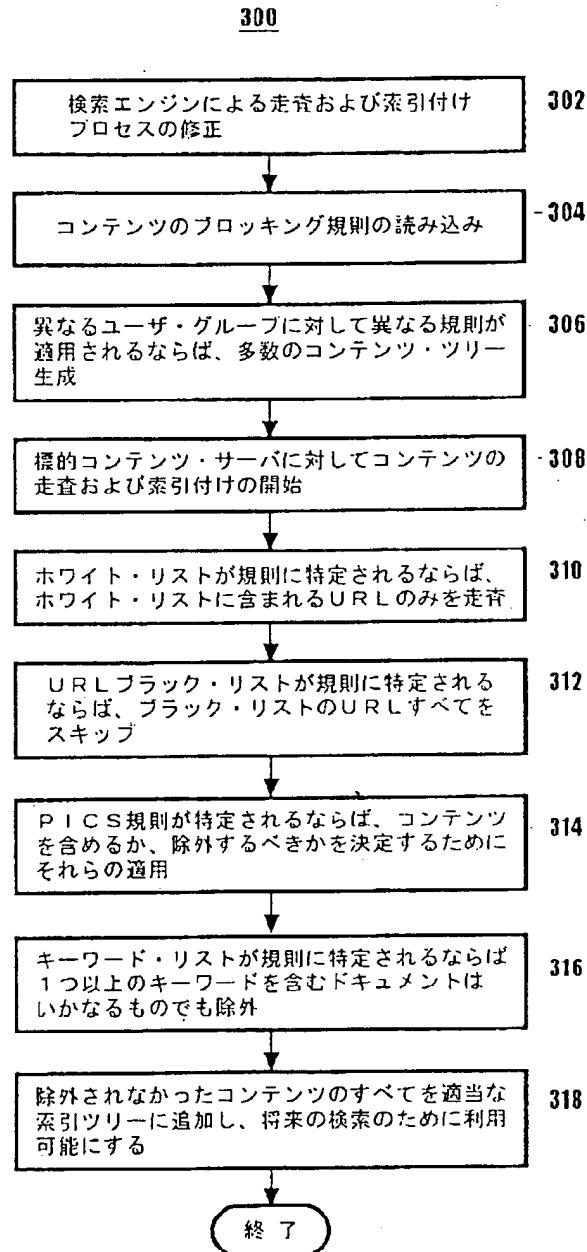
【図2】

200

ユーザー登録ID 203	プロックに対する キーワード 205	PIGS 規則 207	URL ブラック・リスト 209	URL ホワイト・リスト 211
ID	性 詛罵 薬物	ラベル規則1 ラベル規則2 ラベル規則3	URL X URL Y URL Z	URL A URL B URL C

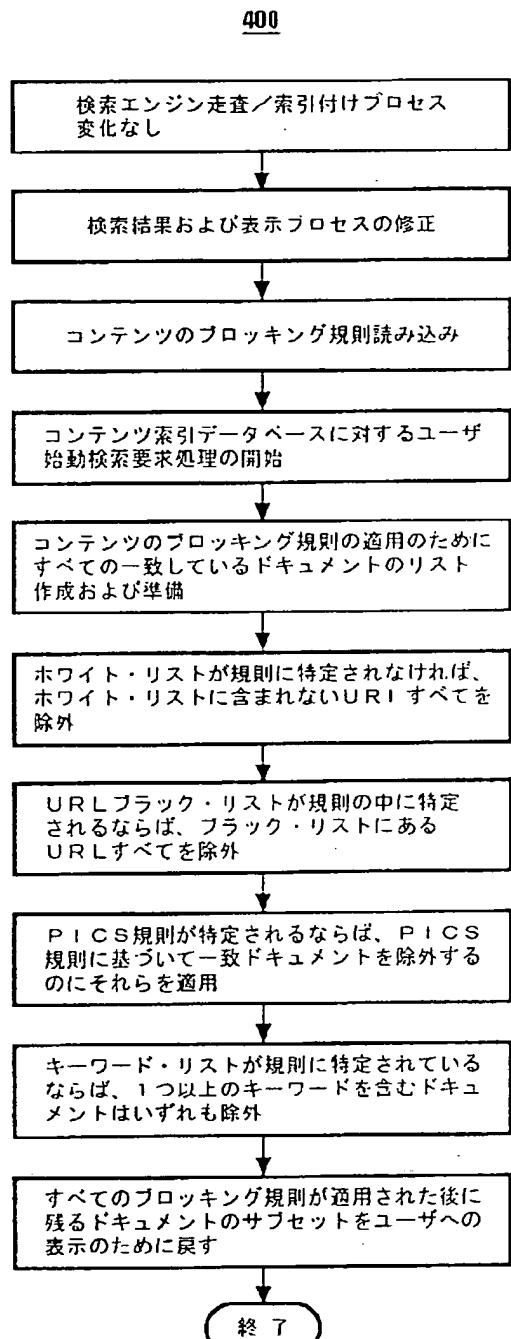
サンプル・コンテンツ・プロッキング規則表

【図3】



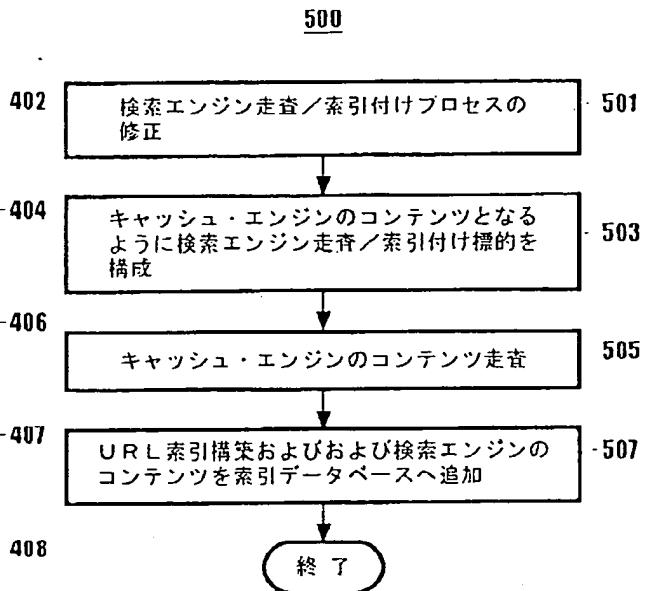
コンテンツ走査／索引付けフェーズの間の実行

【図4】



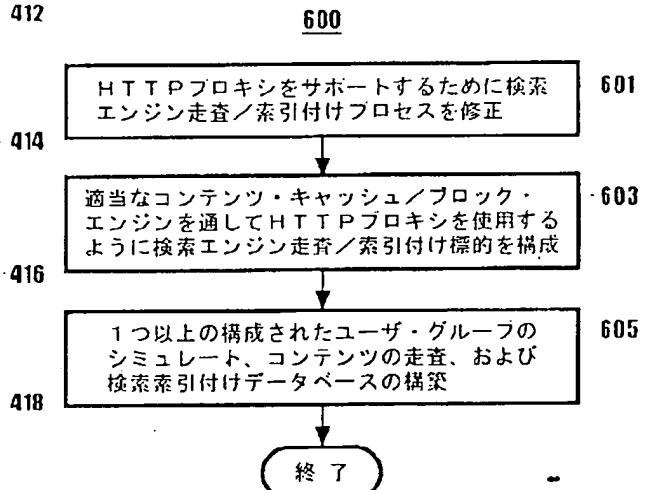
走査結果表示フェーズの間の実行

【図5】



キャッシュ・エンジンのコンテンツから索引付けデータベースの構築

【図6】



キャッシュ・エンジンのコンテンツから索引付けデータベースの構築

フロントページの続き

(51) Int. Cl.⁷

識別記号

F I

テマコード(参考)